Quantization of Nonholonomic systems

Alan Edward Pratt

H.H.Wills Physics Laboratory,
University of Bristol.

A thesis submitted to the University of Bristol
in accordance with the requirements for the degree of
Doctor of Philosophy
in the Faculty of Science.

September 1996

Abstract

In this thesis an investigation is made into the feasibility of using the Feynman path integral formulation to quantize dynamical systems subject to nonholonomic constraints. For these "nonholonomic systems" the classical path does not obey a variational principle in that its action is not stationary with respect to neighbouring paths satisfying the constraints. Consequently, the natural approach of including all paths which satisfy the constraints leads to stationary paths which do not obey the classical equations of motion. Quantum mechanics with unconventional classical motion is the result. The alternative is conventional classical mechanics with no clear generalisation to quantum mechanics. This generalisation is attempted for a simple nonholonomic system. In order to examine propagation over a finite time interval, a model of the constrained system is proposed and investigated within the spirit of path integration.

Acknowledgements

I would like to thank all those who have made this possible. In particular, I would like to thank Dr. J. H. Hannay for suggesting this project and for supervising the work on which this report is based. In particular, I would like to acknowledge the importance of his expertise in the construction of novel optical model systems. Setting up the mathematical framework for such systems and investigating them has occupied me for much of the duration of this project.

I am grateful to the Engineering and Physical Sciences Research Council for financial support.

Author's Declaration

I declare that no part of this thesis has been submitted for a higher degree in this, or any other, university. The research reported herein is the result of my own investigation unless reference is made to the work of others. All research was carried out under the supervision of Dr. J.H.Hannay, at the University of Bristol, between October 1992 and March 1996.

Any opinions expressed in this thesis are my own and not those of the University of Bristol.

Contents

List of Figures

Introduction

Features

This section includes a list of some features of the presentation which may be of interest to the reader and indicates the sections where they are discussed in more detail or "illustrated".

- An account of path integration is given in section 2.4. The reason for using them in this investigation is also discussed (e.g. section 2.2).
- The problem considered here does not appear in the literature. The standard method of quantizing "constrained" systems is discussed in section 2.2 and appendix B. Different approaches which have been applied to a certain type of *holonomic* constraint are discussed in section 4.3.
- The role of models is discussed in section 11.4. The relationships between the models are outlined in the introductions to the relevant chapters (i.e. chapters 6-10). The discussion of approximation is initiated in section 6.1.
- Where mathematical derivations contain lots of simple steps it is considered preferable to describe these in words to avoid unnecessary "clutter" (e.g. equation (10.14)). Where some of the steps are not trivial the calculation is broken into several stages (e.g. equations (10.7)- (10.12)).
- The types of mechanical principle which are relevant to the investigation is discussed in section 1.5.1. Those which *are* relevant have been included in the main text (e.g. section 1.5.2). Others which are not directly relevant are included in appendix A.
- The significance of paraxial optics is discussed in chapter 3. It is appropriate because only nonrelativistic quantum mechanics is considered in this work.
- The goal of the research is stated later in this introductory chapter.

- The conclusions are stated in chapter 11.

Preliminaries

The term "nonholonomic system" refers to a dynamical system subject to a class of constraint not usually considered in quantum mechanics. However such constraints do occur in classical mechanics and presumably also in the quasi-classical limit in some form. We therefore take the most direct approach and attempt to quantize classical systems subject to nonholonomic constraints. We require the quantized system to have the correct classical limit. This prevents the use of standard methods of quantizing a constrained system such as Dirac's procedure [9]. The main aim is to obtain an expression for the propagator using the Feynman Path integral [10], since this offers a direct route from classical to quantum mechanics, but Schrödinger type wave equations are also considered.

Overview

Chapter 1: An introduction to constrained mechanical systems

Chapter 2: A discussion of quantization methods

Chapter 3: An explanation of paraxial optics and its relevance

Chapter 4: An investigation of the possibility of obtaining a wave equation for nonholonomic systems

Chapter 5: An introduction to the special nonholonomic system upon which attention is subsequently focused

Chapter 6: A description of a model for this simple system

Chapter 7: Calculations using the model

Chapter 8: A way to enforce the constraints approximately

Chapter 9: Modifications of the previous approaches

Chapter 10: A version of the model applicable in the nonholonomic limit

Chapter 11: Discussion and Conclusions

Chapters 5–10 deal exclusively with a special case system which is, however, believed to contain the essence of the problem.

With the exception of the first chapter, the presentation is in the form of an edited account of an investigation into the possibility of quantizing mechanical systems subject to nonholonomic constraints. With the possible exception of chapter 4, the chapters follow a continuous line of development.

The goal

The purpose of this thesis (as mentioned in the abstract) is to investigate the possibility of quantizing dynamical systems subject to nonholonomic constraints using the Feynman Path integral formulation. The goal is some form of path integral for a simple nonholonomic system which is sufficiently general to contain the essence of the problem. In Feynman's Path integral formulation of quantum mechanics (discussed in section 2.4), the time evolution of wavefunctions is specified by obtaining the propagator (i.e. the kernel in the integral equation for the wavefunction) in terms of a path integral. Evaluation of the path integral provides sufficient information for the time evolution of a wavepacket to be obtained.

It is of fundamental importance that any expression obtained should have the correct classical limit. If complete success is achieved, a description will have been obtained of the properties of (a set of) quantum systems whose classical limit is a nonholonomic system.

Structure

The structure of this report is illustrated in the the "tree" diagram in figure 0.1. The "axis" down the page represents (roughly) progress towards the goal of the project as stated above (i.e. in the abstract and the preceding section). The direction of the lines are used to distinguish different approaches to the problem. It should be noted that the "tree" in figure 0.1 represents the structure of this "report" rather than the complete history of the research on which it is based. A tree diagram of the latter would require extensive "pruning" to remove "branches" before it resembled figure 0.1.

It is difficult to "score" the relative merits of the three main approaches (branches A, B, C in figure 0.1), they are quite closely related, although the results they provide have differing advantages and disadvantages.

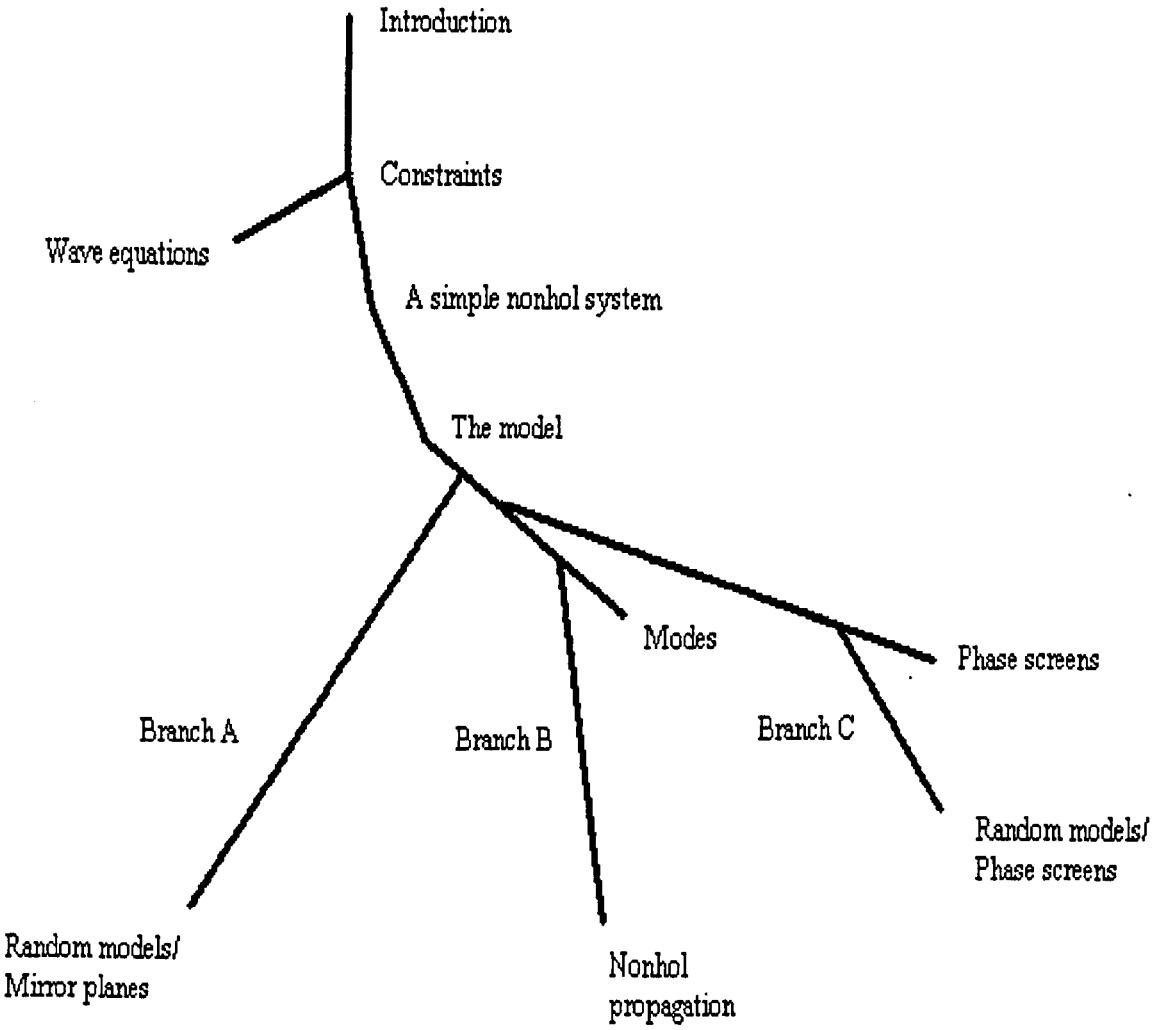


Figure 0.1: Report structure

Chapter 1

Constraints

1.1 Introduction

Constraints limit the motion of a mechanical system, for example the beads of an abacus are constrained to one-dimensional motion by the supporting wires. Constraints are imposed by forces (forces of constraint) but are distinguished from conventional forces in that they are known, or most easily stated, in terms of their effect on the motion of the system.

1.2 Classification of constraints

Constraints can be classified in many ways. A fundamental distinction is between equality constraints and those specified by an inequality, for example a particle confined within a container. Equality constraints may be geometric or kinematic. Constraints are described as geometric if they are expressed by equations involving the position (but not the velocity)

$$f_l(\underline{r}, t) = 0 \tag{1.1}$$

($l = 1, 2, \ldots, m$ where m =number of constraints)
and as kinematic if the equations contain the velocity

$$f_l(\underline{r}, \dot{\underline{r}}, t) = 0 \tag{1.2}$$

Kinematic constraints are integrable if the corresponding system of differential equations is integrable.

Integrable kinematic constraints and geometric constraints, to which they may be reduced, are known as holonomic constraints. Nonholonomic constraints are sometimes taken

to be any constraints which are not holonomic (including "inequality constraints" for example), but, following Hertz (who is generally credited with introducing the term) will be used here to mean specifically non-integrable kinematic constraints.

Constraints are further classified according to whether the equations of constraint contain the time as an explicit variable (rheonomous) or are not explicitly dependent on time (scleronomous).

It is believed that all nonholonomic constraints occurring in nature depend only linearly on the velocity, or equivalently, they may be written as a set of linear differential constraints

$$\sum_k a_{lk} dq_k + a_{lt} dt = 0 \tag{1.3}$$

for generalized coordinates q_k, $k = 1, \ldots, n$

such a form of the equations also includes holonomic constraints, i.e.

$$\sum_k \frac{\partial f_l}{\partial q_k} dq_k + \frac{\partial f_l}{\partial t} dt = 0 \tag{1.4}$$

for holonomic constraints $f_l(\underline{q}, t) = 0$

and is known as the Pfaffian form of the constraint equations.

A constraint equation of the form (1.3) is called catastatic when $a_l = 0$ otherwise it is called acatastatic.

1.3 Examples of constrained systems

An example of a dynamical system with a *holonomic* constraint is a frictionless bead on a horizontal circular wire. The two cartesian coordinates which would be required to locate the bead in the horizontal plane reduce to a single angle coordinate. This system is trivially quantizable — "the rotor".

Similarly, a vertical disc of radius r rolling without slipping along a horizontal line is a dynamical system with a holonomic constraint: the velocity $\dot{x}$ and the angular velocity $\dot{\theta}$ of the disc are linked by $\dot{x} = r\dot{\theta}$. By integration, therefore, the angle θ and the contact position x of the disc are linked, and one can be discarded. This system is also straightforward to quantize.

In contrast, a similar system with a *non-holonomic* constraint is a disc whose radius is a (prescribed) function of time rolling without slipping on a horizontal line. Now $\dot{x} = r(t)\dot{\theta}$ which cannot be generally integrated to link the x and θ coordinates — both are needed. [The space x, θ, t is no longer filled with a stack of sheets $f(x, \theta, t) = const.$ to which the

motion is confined]. This is probably the simplest type of non-holonomic system, and will be considered later with regard to quantization.

Two better known examples of non-holonomic classical systems are a vertical skate on ice, and a ball rolling on a perfectly rough surface. In the first case the nonholonomic constraint is the requirement that the velocity of the skate in the direction perpendicular to the plane of its blade is zero. In the second case the velocity of the point of contact must vanish. The fact that the constraints cannot be integrated to obtain relations between the coordinates may be illustrated in the second case by rolling the ball from a certain initial position along two different paths so that the two final positions of the point of contact coincide. Generally, the final orientation of the ball is found to be different for each path.

1.4 A geometrical picture

If a two-freedom dynamical system is represented as a point in three dimensional space-time then constraints have a simple geometrical interpretation (this is also true for a three-freedom system with a time independent "catastatic" constraint represented by a point in 3D space, although such a case is not considered in later chapters within the main text, it effectively reduces to the two freedom system described if an additional constraint $\dot{z} = 1$ is included). A kinematic constraint defines a field of tangent planes or "planelets" in the three dimensional space-time. The constraint is that the tangent to the path (world-line) of the particle "lies within" the infinitesimally small plane defined at the particles current position.

If the constraint is holonomic, then the planelets fit together to form surfaces. Thus any possible trajectory lies within a surface. So a holonomic constraint restricts the motion to a 2D subspace of the original 3D space-time.

By contrast, for a nonholonomic constraint, the planelets do not form a surface. Any two points may be joined by a path, not obeying any equations of motion, but at least satisfying the constraints. So the whole space is "geometrically accessible".

In 3D space the condition for the field of planelets (associated with a "catastatic" constraint) to be holonomic is that

$$\underline{N}.(\underline{\nabla} \times \underline{N}) = 0 \tag{1.5}$$

where $\underline{N} = (n_x(x,y,z), n_y(x,y,z), n_z(x,y,z))$ is the normal vector to the planelet at the specified point.

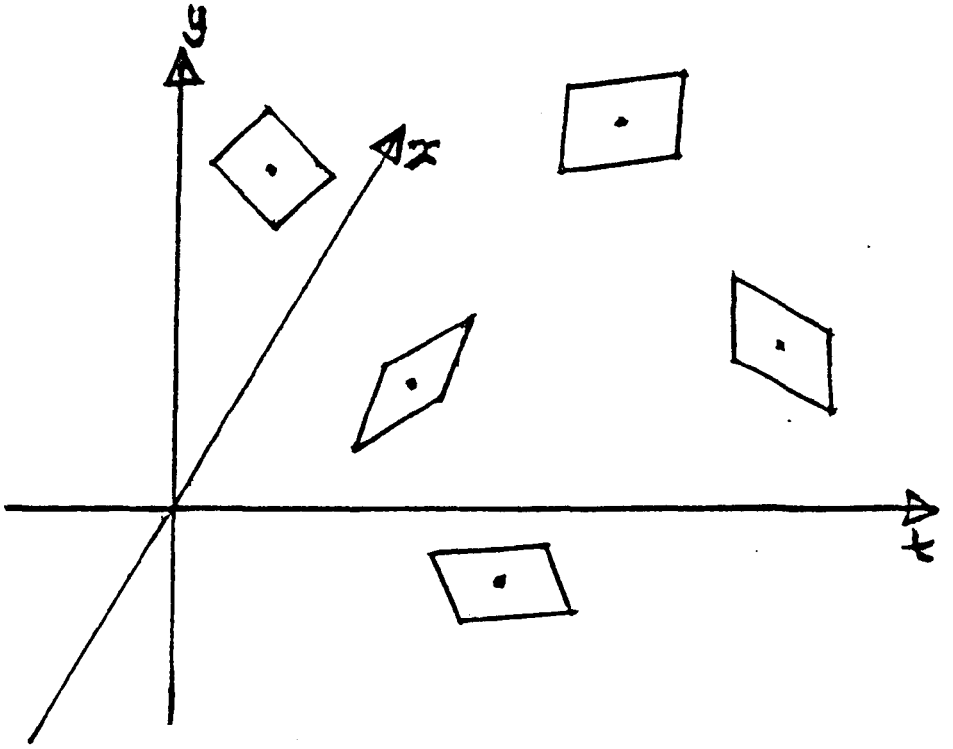


Figure 1.1: A field of "infinitesimal" tangent planes

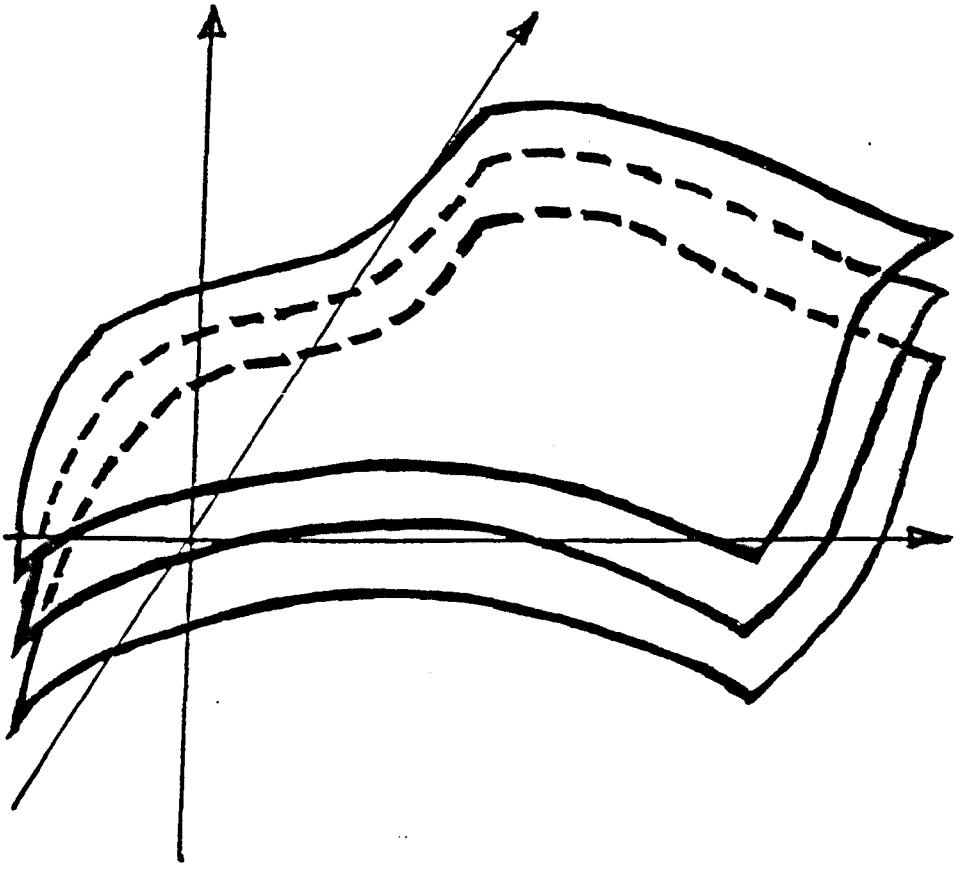


Figure 1.2: "Constraint surfaces" for holonomic constraints

(for 3D space-time one simply replaces z by t in the formula) This is a special case of the Frobenius condition [3]

$$\tilde{\omega} \wedge \tilde{d\omega} = 0 \tag{1.6}$$

for the integrability of a field of hyperplanes

$$\tilde{\omega} = 0 \tag{1.7}$$

where $\tilde{\omega}$ is a 1-form.

1.5 Mechanics of constrained systems

1.5.1 Mechanical principles

A way of classifying the principles of classical mechanics is to split them into two groups depending upon whether they can be derived from a principle of stationary action in all cases. The proviso "in all cases" means cases with nonholonomic constraints must be included. This is important because in unconstrained mechanics (and also when holonomic constraints are present) the distinction can largely be removed by suitable manipulation of the equations. With this proviso the groups are:

class A: equivalent to a principle of stationary action

class B: all other accepted comprehensive mechanical principles

where "action" is taken to mean the integral over time of some well defined quantity (i.e. sufficiently general to allow constraints to be included) and the use of "stationary" is the same as in the standard calculus of variations. The reason for making this distinction is that the Feynman Path integral formulation is based on a principle of stationary action [10, 32]. The standard Feynman path integral formulation can only be applied if there is a variational principle which gives the correct equations of motion for a classical system subject to nonholonomic constraints. Otherwise some sort of generalization of the Feynman formulation must be attempted.

Examples of procedures for obtaining the correct nonholonomic equations of motion include [2] D'Alembert's principle, Gauss's principle and the Gibbs-Appell equations (appendix A). All of these fall into "class B". The question is: can a principle found in "class A" give the correct nonholonomic equations of motion? The answer is no [31, 18, 29]. When the principle of stationary action is applied to a constrained system the result is

"vakonomic mechanics". When the constraints are integrable (holonomic) this reduces to ordinary holonomic mechanics.

1.5.2 Vakonomic mechanics

For holonomic constraints, $g_l(\underline{q},t) = 0$, there is a variational principle which gives the correct equations of motion: the principle of stationary action is applied to the subspace of the original space defined by the constraints. This may be achieved using multipliers in the standard way: the unconstrained Lagrangian $L(\underline{q},\underline{\dot{q}},t)$ is replaced by $L(\underline{q},\underline{\dot{q}},t) + \sum_{l=1}^{m} \lambda_l g_l(\underline{q},t)$ (the "modified Lagrangian") and the multipliers λ_l are treated as independent coordinates. The resulting Euler-Lagrange equations give the equations of motion (from variations with respect to $\underline{r}(t)$)

$$-\frac{d}{dt}\frac{\partial L}{\partial \underline{\dot{q}}} + \frac{\partial L}{\partial \underline{q}} = -\sum_l \lambda_l \frac{\partial g_l}{\partial \underline{q}} \tag{1.8}$$

and the constraint equations (from variations with respect to $\lambda(t)$)

$$g_l(\underline{q},t) = 0 \tag{1.9}$$

allowing the "elimination" of the multipliers (i.e. solving for λ so that $\lambda = \lambda(\underline{q},\underline{\dot{q}},t)$)

If the holonomic constraint equations are written in the kinematic form

$$\dot{g}_l(\underline{q},t) = 0 \tag{1.10}$$

then the resulting equations of motion are similar but λ_l is replaced by $-\dot{\lambda}_l$. This difference can be removed simply by defining new multipliers $\mu_l = -\dot{\lambda}_l$. The multipliers are obtained by solving algebraic equations.

Vakonomic "mechanics" consists in applying this variational procedure regardless of whether the constraints are holonomic or not, i.e. taking it as an "axiom" (hence "vak" from mechanics of variational axiomatic kind [2]). So the "modified Lagrangian" is taken to be

$$L(\underline{q},\underline{\dot{q}},t) + \sum_{l=1}^{m} \lambda_l f_l(\underline{\dot{q}},\underline{q},t)$$

where the general constraints $f_l(\underline{\dot{q}},\underline{q},t) = 0$ may be holonomic or nonholonomic. Applying the principle of stationary action (this is equivalent in space-time to asking for the "shortest" path amongst those satisfying the constraints) produces the equations of motion

$$-\frac{d}{dt}\frac{\partial L}{\partial \underline{\dot{q}}} + \frac{\partial L}{\partial \underline{q}} = \sum_l \dot{\lambda}_l \frac{\partial f_l}{\partial \underline{\dot{q}}} - \sum_l \lambda_l \left[-\frac{d}{dt}\frac{\partial f_l}{\partial \underline{\dot{q}}} + \frac{\partial f_l}{\partial \underline{q}} \right] \tag{1.11}$$

There are three cases to consider.

The first case occurs when the constraints are holonomic, with no dependence on velocity, i.e. equations (1.9), so that

$$f_l(\underline{\dot{q}}, \underline{q}, t) = g_l(\underline{q}, t) \tag{1.12}$$

In this case equations (1.11) reduce to equations (1.8), exactly as expected.

The second case occurs when the constraints are again holonomic (equations (1.9)) but are now written in the kinematic (differentiated) form (1.10) so that

$$\begin{aligned} f_l(\underline{\dot{q}}, \underline{q}, t) &= \dot{g}_l(\underline{q}, t) \\ &= \frac{\partial g_l}{\partial t} + \underline{\dot{q}} . \frac{\partial g_l}{\partial \underline{q}} \end{aligned} \tag{1.13}$$

In this case

$$\begin{aligned} -\frac{d}{dt}\frac{\partial f_l}{\partial \underline{\dot{q}}} + \frac{\partial f_l}{\partial \underline{q}} &= -\frac{d}{dt}\frac{\partial g_l}{\partial \underline{q}} + \frac{\partial}{\partial \underline{q}}\dot{g}_l \\ &= 0 \end{aligned} \tag{1.14}$$

so the equations of motion become

$$-\frac{d}{dt}\frac{\partial L}{\partial \underline{\dot{q}}} + \frac{\partial L}{\partial \underline{q}} = \sum_l \dot{\lambda}_l \frac{\partial g_l}{\partial \underline{\dot{q}}} \tag{1.15}$$

Again, exactly as described (for constraints (1.10)).

The third case is when the constraints are nonholonomic — i.e.

$$\textit{nonintegrable} \;\; f_l(\underline{\dot{q}}, \underline{q}, t) = 0 \tag{1.16}$$

In this case it is not in general possible to simplify equations (1.11). These equations of motion are inconsistent with the accepted equations describing nonholonomic mechanical systems due to the presence of the terms containing the square brackets.

The presence of both λ and $\dot{\lambda}$ in (1.11) means that the equations to determine the multipliers are differential equations rather than algebraic equations and so constants of integration associated with the multipliers are now required. These are arbitrary, choosing them suitably allows any final point to be reached from a given initial point. The name (from [2]) "vakonomic mechanics" (vak) will be used for the mathematical formalism based on equations (1.11) and (1.16), derived by comparing paths which satisfy the constraints and requiring the action to be stationary (using the standard calculus of variations techniques). It does not agree with the experimentally observed [23] nonholonomic classical mechanics — "ordinary mechanics" (ord). It must, therefore, be rejected as unphysical for the purposes of classical mechanics.

1.5.3 Comparison

Before comparing vakonomic and ordinary mechanics, it is worth noting a possible source of confusion: in some treatments the Euler-Lagrange equations are derived for the unconstrained system and are then modified to correctly take into account any constraints present, the resulting equations are

$$-\frac{d}{dt}\frac{\partial L}{\partial \underline{\dot{q}}} + \frac{\partial L}{\partial \underline{q}} = \sum_l \Lambda_l \frac{\partial f_l}{\partial \underline{\dot{q}}} \tag{1.17}$$

where Λ_l are refered to as "multipliers".
Solving these together with the constraint equations (1.16) gives the physical motion. It is important to realise that such equations giving the correct nonholonomic equations of motion cannot be obtained directly from a principle of stationary action (i.e. they fall into class B). These correct (ord) equations of motion (equations (1.17)) do not include derivatives of the "multipliers", in contrast to the vak equations.

In vak mechanics constants of integration associated with the multipliers are required when the constraints are nonholonomic, in ord mechanics this is not the case. So in order to determine the motion in vak mechanics with nonholonomic constraints, one must supply more information than is required for conventional (ord) mechanics. In the holonomic case this problem does not arise (section 1.5.4).

A final point is defined to be "dynamically accessible" from a given initial point if it may be reached from the initial point by a path satisfying the equations of motion. In vakonomic mechanics with nonholonomic constraints there is *no* reduction of the dimension of space which is dynamically accessible from any given initial point, despite the presence of constraints. In ordinary nonholonomic (ord) mechanics the dimension of space dynamically accessible from a given initial point *is* reduced. For example, in 3D space time with one constraint this means that the initial position (2 coordinates) and velocity (1 number since the constraint must be satisfied) are sufficient to determine the motion for ord mechanics and a curve in the final plane is dynamically accessible (by taking all possible values of initial velocity) from a point in the initial plane (this is like a contact transformation [20, 36]). Specifying the initial position in vak mechanics still allows any point in the final plane to be reached, to determine the motion requires two more numbers such as the final coordinates.

So vak mechanics is not identical to ord mechanics but amongst the vak paths are a subset which have the same final points as ord mechanical paths. The question is: if these final points are specified in the vak formalism, are the resulting vak paths the same as the

ordinary mechanical paths? i.e. does vak mechanics "contain" ord mechanics? The answer is no (the "routes" differ), it is not possible to remove the extra term in the vak equations of motion.

In terms of forces in (potential free) 3D space-time, the difference between the vak and ord equations of motion (for a single constraint) is that for the ord case the only forces acting are those required to ensure that the system satisfies the constraint. The force acts in a direction perpendicular to the relevant planelet. The component of the force in this direction is determined by taking the time derivative of the constraint equation, this is a valid equation since the constraint must be satisfied for all values of time. The ord prescription is to take all other components of the force to be zero. In the vak case the force has a component parallel to the planelet, causing the path to curve "within the planelet" (as shown in figure 1.3). The vak prescription is for a path of stationary length (compared to other paths satisfying the constraints). Consequently, the components of the force not determined by the constraint take whatever values are needed to meet this requirement. The greater freedom in the vak case allows any final point to be reached from the given initial point. Even when the paths go between the same points the vak path will still "curve within the planelets" and the ord path will not. As noted by Hertz [17] (for the zero potential case), the vak path is the shortest ("of stationary length" strictly speaking) and the ordinary mechanical path the straightest, consistent with the constraints. In general these will not coincide.

1.5.4 Holonomic case

As explained in section 1.4 (for 3D space-time) a kinematic holonomic constraint defines a "stack of surfaces" within the space. If the kinematic constraint is integrated and the constant of integration is specified to give a single geometric constraint, then a single surface is defined.

It is now desired to consider, in addition, the dynamics. For kinematic holonomic systems the dynamics is defined by the principle of stationary action within the subspaces ("surfaces"). Specifying the initial position determines which surface within the stack the motion takes place on (for a geometric holonomic constraint consistency is required) and also the initial position on the surface. If the initial velocity is consistent with motion on the surface then one can "do mechanics" on the subspace. However, when the constraints are nonholonomic, no such surfaces are formed.

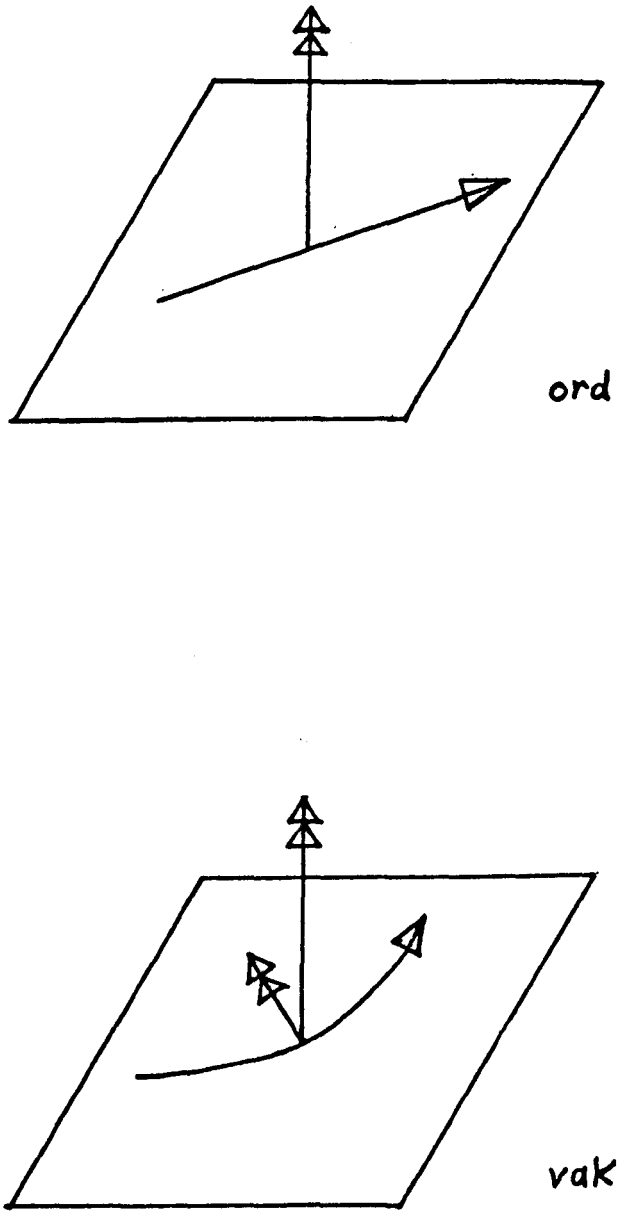


Figure 1.3: "Magnified planelets" for the "ord" and "vak" cases

1.5.5 The classical fan

An idea introduced in section 1.5.3 will be used again later: in 3D space-time with one constraint, the initial position is specified and the initial velocity is allowed to take all values consistent with the constraints (1 parameter). The set of paths in space-time obeying the correct equations of motion (i.e. the ord equations in the nonholonomic case) with these initial conditions will be called the "classical fan". The final time defines a $t = constant$ plane in space-time. The intersection of the classical fan with this plane produces a curve (the "intersection curve"). If the constraints are holonomic, then the classical fan always lies in a space-time surface formed by the constraints, and the "intersection curve" coincides with a constraint "surface" at the final time (which is a curve in 2D). In two (space) dimensions it is always possible to construct a set of curves (in any $t = constant$ plane) by joining the infinitesimal line segments representing the constraints. This is special to the case of two (space) dimensions and does not mean that it is possible to construct constraint surfaces in the 3D space-time (unless the constraints are *holonomic*, of course). In the nonholonomic case the "intersection curve" will not, in general, coincide with any curve constructed in this way (figure 1.4).

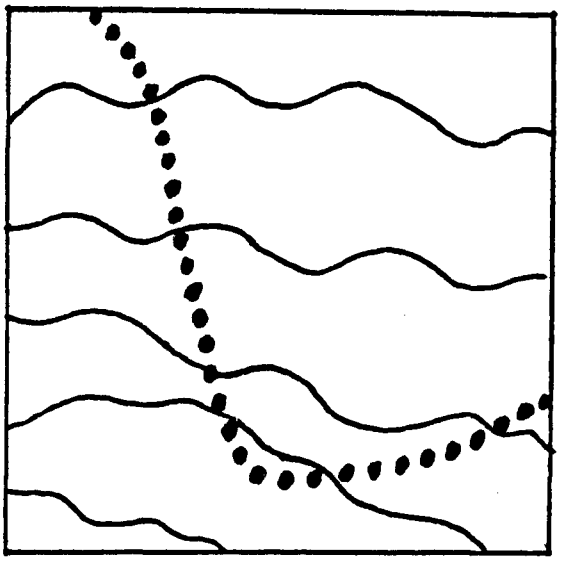
1.6 Constrained Hamiltonian systems

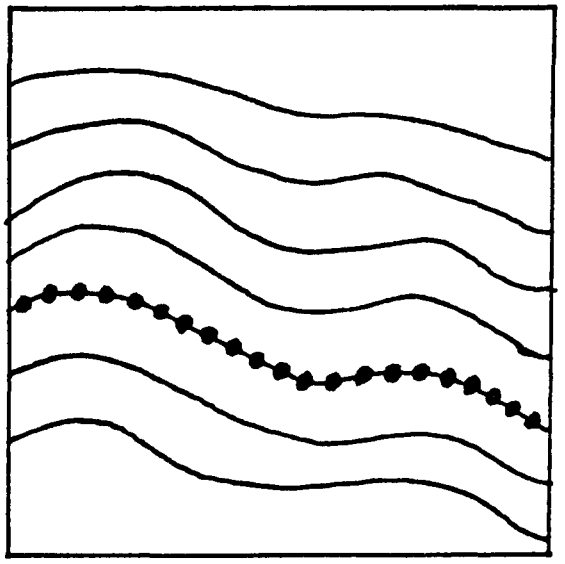
There is a method for quantizing "constrained" systems which follows from the work of Dirac [9]. Consequently the question arises as to whether this can be applied to nonholonomic systems. There are two parts to the process: the first is to obtain the classical Hamiltonian dynamics of the "constrained" system; the second is to quantize this using the canonical quantization procedure. The first part of this process is considered here.

In this section the term "constrained" takes a different meaning from the one that it has in the rest of this chapter. In this section "constrained" takes on the "technical" meaning that it has in the field of "constrained dynamics". In "constrained dynamics" a (N degree of freedom) non-relativistic system with Lagrangian $L(q^i, \dot{q}^i, t)$ is said to be "constrained" if the matrix

$$W_{ik} = \frac{\partial^2 L}{\partial \dot{q}^i \partial \dot{q}^k} \tag{1.18}$$

has zero determinant i.e. $W = \det W_{ik} = 0$. The lagrangian is then "singular". If $W \neq 0$ it is "regular". Systems with $W = 0$ are called "singular Lagrangian systems", "constrained Hamiltonian systems" or "degenerate systems". In classical mechanics this





•••• "intersection curve"
—— joined up "constraint line segments"

Figure 1.4: "Intersection curves" for holonomic and nonholonomic cases

nomenclature is potentially confusing, although constraints in the usual sense (holonomic and nonholonomic) do make a Lagrangian singular if Lagrange multipliers are considered as dynamical variables.

The most basic question to be asked is whether the classical Hamiltonian equations resulting from applying the techniques of "constrained dynamics" to a nonholonomic system agree with the accepted classical mechanical equations of motion for nonholonomic systems. The answer is that they do not: more details are given in appendix B. The equations are correct for holonomic constraints but not for nonholonomic constraints. In fact they agree with the equations of vakonomic "mechanics" (appendix B). This is not unexpected since in both cases Lagrange multipliers are treated as dynamical variables.

The conclusion is that the Dirac quantization procedure is not suitable for the quantization of nonholonomic systems. The requirement that the quantized system should have the correct classical limit will not be met if the Dirac procedure is used when nonholonomic constraints are present. This is because quantization is applied to the "wrong" nonholonomic classical system.

1.7 Summary

The main objectives of this chapter were:

1. To provide an introduction to constraints in classical mechanics.

2. To introduce a geometrical picture suitable for some cases of interest, including a special case that will be important in the following chapters (e.g. section 5.2).

3. To show that the principle of stationary action does not give the correct equations of motion of classical mechanics when nonholonomic constraints are present. This is important because it is just this principle which is required for the standard Feynman path integral quantization. Quantization is discussed in later chapters (i.e. chapter 2 and section 5.4).

4. To indicate that the approach to classical mechanics with constraints based on Dirac's method does not give the correct results when the constraints are nonholonomic.

Chapter 2

Quantization

2.1 Introduction

In this chapter methods of quantization are considered. It is explained why the Feynman (configuration space) path integral is chosen as the method of quantization. An introduction to path integrals is included.

2.2 Canonical quantization

Canonical quantization is the longest established method of quantizing a dynamical system. It provides a set of rules for passing from Hamilton's dynamics to quantum dynamics, by making the coordinates and momenta into linear operators.

The Dirac method for quantizing "constrained" dynamical systems uses canonical quantization. The first stage of this procedure is to pass from the Lagrangian to the Hamiltonian description of the classical dynamics. This puts the system in a suitable form to apply the second stage, which is to pass to quantum dynamics using the canonical quantization rules.

As explained in section 1.6 there are problems with the first stage of the Dirac procedure when nonholonomic constraints are present. Consequently, it is desirable to avoid the first stage (passage from Lagrangian to Hamiltonian dynamics). This means abandoning canonical quantization (which is applied to the Hamiltonian description). The method used to quantize systems starting directly from the Lagrangian description is Feynman's path integral quantization procedure. Consequently, this is the method upon which the main part of his thesis is based. In fact, a type of "path integral" quantization has been used in developments of the Dirac procedure but this involves phase space functional in-

tegrals (rather than the position space variety). This is an important distinction because phase-space functional integrals require canonical momenta to be (well) defined. The problems described here are not a consequence of operator ordering ambiguities in canonical quantization as such ambiguities also appear in path integral quantization as questions of where to evaluate functions in the Lagrangian [32].

It is worth mentioning (as an aside) that even for systems whose constraints are not nonholonomic, things are not straight-forward because there isn't a unique rule for the canonical quantization of constrained systems [33].

2.3 Path integral quantization

In order to quantize an unconstrained dynamical system using the Feynman path integral formulation, the classical action for all possible paths between the two end points is required and the action must be stationary on the classical dynamical path [10].

For a system subject to holonomic constraints this prescription should be applied in the subspace of the original space defined by the constraints, although complications arise due to curvature of the subspace.

For a system subject to nonholonomic constraints the analogous procedure is to include all paths satisfying the constraints. As described in appendix C, this is the most obvious way to proceed, however it is equivalent to the quantization of vakonomic "mechanics". From previous study of vakonomic mechanics (section 1.5.2) we know what undesirable features to expect when the constraints are nonholonomic. The stationary paths do not obey the nonholonomic classical mechanical equations of motion. The correct classical path does not obey a variational principle in that its action is not stationary with respect to neighbouring paths satisfying the constraints so it will not be recovered in the classical limit. This leaves a choice between:

- Quantum mechanics with unconventional classical (i.e. stationary) motion — the vak case
- Classical mechanics (conventional) with no clear generalization to quantum mechanics — the "ord" case

2.4 Path integrals

2.4.1 Introduction

It is desired to use the concepts of path integration to explore the quantum mechanics of a novel class of systems. Consequently, the path integral formulation is introduced first and then shown to be equivalent to standard Schrödinger quantum mechanics, as in the original work of Feynman [11], rather than using standard quantum mechanics to justify the construction of the path integral (e.g. [32]).

2.4.2 Construction

In quantum mechanics the fundamental quantities are probability amplitudes, φ_{ab}. If $P(a,b)$ is the probability to go from a state a to a state b, then the relation to the corresponding probability amplitude is given by

$$P(a,b) = |\varphi_{ab}|^2 \tag{2.1}$$

The "composition" rule for probability amplitudes depending on two states is

$$\varphi_{ab} = \sum_c \varphi_{ac}\varphi_{cb} \tag{2.2}$$

where the sum is over all possible states c.

This relation may be used in the construction of the of a sum over all paths. Considering the 1D case, the initial state is x_a at time t_a and the final state is x_b at time t_b. Between t_a and t_b, a set of values of time (t_i for $i = 1,\ldots,N-1$) is taken, with an interval ϵ between consecutive values (i.e. $\epsilon = t_{i+1} - t_i$ and $N\epsilon = t_b - t_a$). At each t_i a point x_i is selected. A ("polygonal") path is constructed by connecting all such points with straight lines (figure 2.1). It is possible to sum over all paths constructed in this way by taking a multiple integral over all values of x_i for i from 1 to $N-1$. This yields an expression for the amplitude ("kernel") for propagation from $(x_0,t_0) = (x_a,t_a)$ to $(x_N,t_N) = (x_b,t_b)$, i.e. (given that $t_b > t_a$)

$$K(x_b,t_b;x_a,t_a) \sim \int\int\cdots\int \phi_N[x(t)]\,dx_1dx_2\cdots dx_{N-1} \tag{2.3}$$

Making ϵ smaller gives a more representative sample of the complete set of all possible paths between the fixed end points. Also, sections of the classical orbit could be used between consecutive points [11] instead of straight lines.

The contribution $\phi_N[x(t)]$ from each path may be obtained using

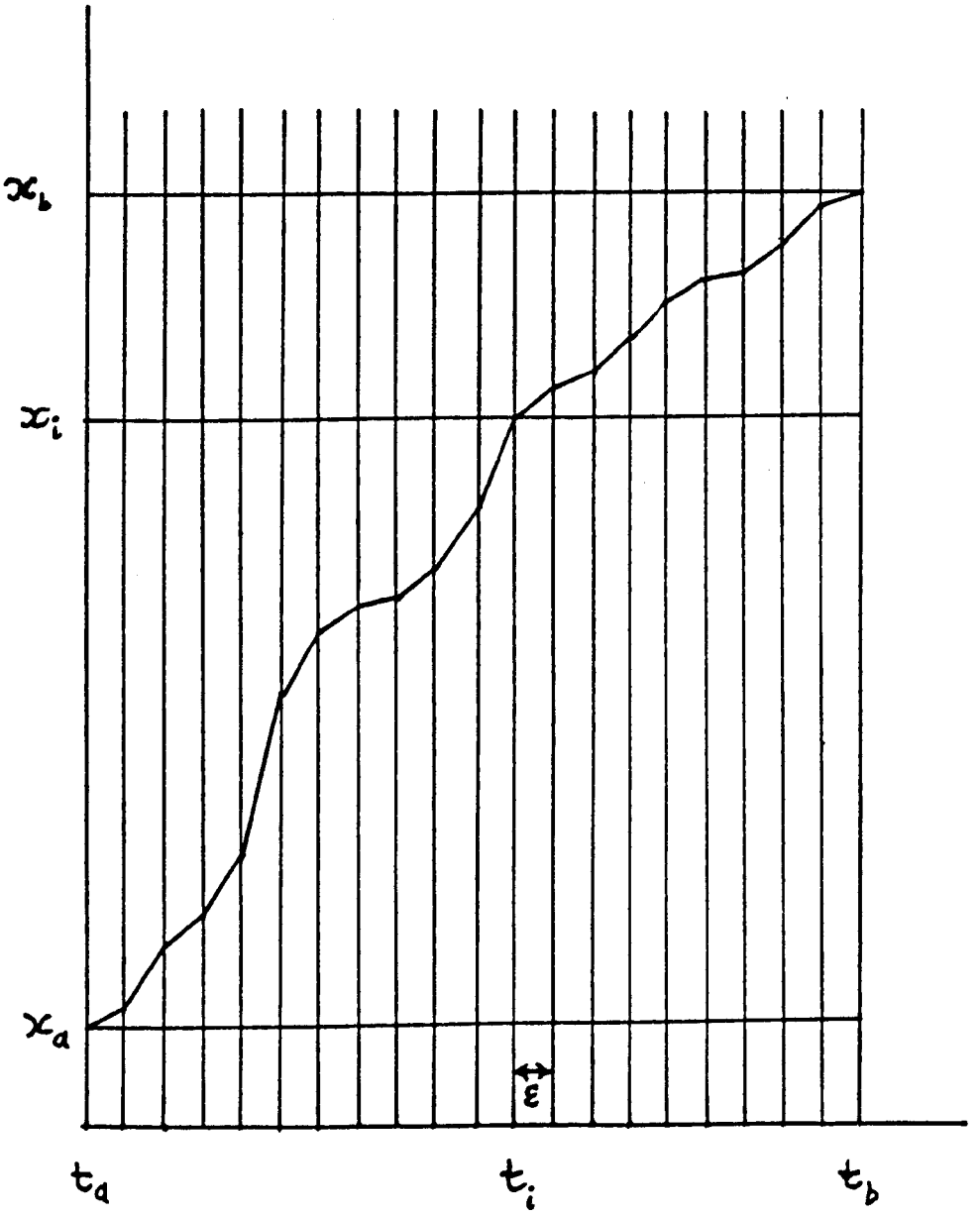


Figure 2.1: A polygonal path

1. The expression

$$K(x_{i+1},t_{i+1};x_i,t_i) = \frac{1}{A}\exp\left[\frac{i\epsilon}{\hbar}L\left(\frac{x_{i+1}-x_i}{\epsilon},\frac{x_{i+1}+x_i}{2},\frac{t_{i+1}+t_i}{2}\right)\right]+O\left(\epsilon^2\right) \quad \text{as } \epsilon \to 0 \tag{2.4}$$

for the (normalised) kernel when $t_{i+1} - t_i = \epsilon$ is an infinitesimal time interval. ($L(\dot{x},x,t)$ is the Lagrangian)

2. The rule that amplitudes for events occurring in succession in time multiply.

The result is

$$\begin{aligned}\phi_N[x(t)] &= \prod_{i=0}^{N-1} K(x_{i+1},t_{i+1};x_i,t_i) \\ &= \frac{1}{A^{N-1}}e^{\frac{i}{\hbar}S_N}\end{aligned} \tag{2.5}$$

where

$$S_N = \sum_{i=0}^{N-1}\epsilon L\left(\frac{x_{i+1}-x_i}{\epsilon},\frac{x_{i+1}+x_i}{2},\frac{t_{i+1}+t_i}{2}\right) \tag{2.6}$$

is (a good approximation to) the action for the path.

Using these results (and noting that $N \to \infty \Rightarrow \epsilon \to 0$) gives (for $t_b > t_a$)

$$K(x_b,t_b;x_a,t_a) = \lim_{N\to\infty}\frac{1}{A}\int\int\cdots\int\exp\left(\frac{i}{\hbar}S_N\right)\frac{dx_1}{A}\frac{dx_2}{A}\cdots\frac{dx_{N-1}}{A} \tag{2.7}$$

2.4.3 The concept

In certain special cases a particular way of constructing the path may prove disadvantageous: for example the construction described above and illustrated in figure 2.1 gives discontinuities in the velocity at the points (x_i,t_i). The fact that the acceleration is infinite at these points could cause problems if the Lagrangian depended on the acceleration. However, in such cases the "substitution"

$$\ddot{x} = \frac{1}{\epsilon^2}(x_{i+1} - 2x_i + x_{i-1}) \tag{2.8}$$

is usually adequate. This is an illustration of the generality of the concept of a sum over all paths and suggests the use of a notation such as

$$K(x_b,t_b;x_a,t_a) = \int_{x_a,t_a}^{x_b,t_b} e^{\frac{i}{\hbar}S}d^\infty x(t) \tag{2.9}$$

which is independent of a particular definition. The expression equation (2.9) is valid for $t_b > t_a$. It is conventional to define $K(x_b,t_b;x_a,t_a)$ to be zero for $t_b < t_a$. For the remaining case (i.e. $t_b = t_a$), the result $K(x_b,t_b;x_a,t_a) \to \delta(x_b - x_a)$ as $t_b \to t_a+$) may be invoked. Unless stated otherwise, subsequent results in this thesis will be for the case "$t_b > t_a$".

2.4.4 The Schrödinger equation

The path integral formulation of quantum mechanics is verified by propagating a wave function at time t, $\psi(x,t)$, to time $t+\epsilon$ using the path integral propagator and showing that this evolution is the same as that given by the Schrödinger equation. This is achieved by applying the general (1D) equation (for $t_2 > t_1$)

$$\psi(x_2,t_2) = \int_{-\infty}^{\infty} K(x_2,t_2;x_1,t_1)\psi(x_1,t_1)dx_1 \tag{2.10}$$

to the special case with the time t_2 differing only by an infinitesimal interval ϵ from t_1 (so $t_1 = t$, $t_2 = t+\epsilon$). In this case the propagator is given by equation (2.4). For the case of a particle in 1D subject to a scalar potential the Lagrangian is $L = \frac{1}{2}m\dot{x}^2 - V(x,t)$. The result of making these substitutions in equation (2.10) is (with $x_2 = x$)

$$\psi(x,t+\epsilon) = \int_{-\infty}^{\infty} \frac{1}{A} e^{\frac{im\eta^2}{2\hbar\epsilon}} e^{-\frac{i\epsilon}{\hbar}V(x+\frac{\eta}{2},t)} \psi(x+\eta,t)d\eta \tag{2.11}$$

The substitution $x_1 = x_2 + \eta$ used in this equation is suggested by the method of stationary phase: the first exponential oscillates very rapidly unless $\frac{m\eta^2}{\hbar\epsilon}$ is small. Consequently most of the integral is contributed by values of η of order $\sqrt{\frac{\epsilon\hbar}{m}}$. This suggests making the expansion

$$\psi(x+\eta,t) = \psi(x,t) + \eta\frac{\partial\psi}{\partial x} + \frac{1}{2}\eta^2\frac{\partial^2\psi}{\partial x^2} + \ldots \tag{2.12}$$

in addition to

$$\psi(x,t+\epsilon) = \psi(x,t) + \epsilon\frac{\partial\psi}{\partial t} + \ldots \tag{2.13}$$

and

$$e^{-\frac{i\epsilon}{\hbar}V} = 1 - \frac{i\epsilon}{\hbar}V + \ldots \tag{2.14}$$

The requirement that both sides of equation (2.11) agree in the limit $\epsilon \to 0$ determines A, i.e.

$$\begin{aligned} A &= \int_{-\infty}^{\infty} e^{\frac{im\eta^2}{2\hbar\epsilon}} d\eta \\ &= \left(\frac{2\pi i\hbar\epsilon}{m}\right)^{\frac{1}{2}} \end{aligned} \tag{2.15}$$

Performing further Gaussian integrals to obtain terms of order ϵ yields

$$\psi + \epsilon\frac{\partial\psi}{\partial t} = \psi - \frac{i\epsilon}{\hbar}V\psi - \frac{\hbar\epsilon}{2im}\frac{\partial^2\psi}{\partial x^2} + o(\epsilon) \qquad as\ \epsilon \to 0 \tag{2.16}$$

So ψ satisfies

$$-\frac{\hbar^2}{2m}\frac{\partial^2\psi}{\partial x^2} + V\psi = i\hbar\frac{\partial\psi}{\partial t} \tag{2.17}$$

which is the Schrödinger equation, as required.

The generalization to three dimensions is straightforward. Also, a magnetic field may be included provided the vector potential, $\underline{A}(\underline{x})$, is evaluated at the midpoint $\frac{1}{2}(\underline{x}_i + \underline{x}_{i+1})$ or the quantity $\frac{1}{2}[\underline{A}(\underline{x}_i) + \underline{A}(\underline{x}_{i+1})]$ is used [32].

2.4.5 The classical limit

The classical limit "$\hbar \to 0$" means that $\hbar$ is small compared to a typical action. The effect of taking this limit in the path integral (sum over paths) can be illustrated by considering a path and then making a small (on the classical scale) change to it. This small change will, in general, produce a large change in the phase, $\frac{S}{\hbar}$, associated with the path. The contribution of a path is proportional to $e^{\frac{i}{\hbar}S}$ which oscillates rapidly as the path is changed. Consequently, if a path is chosen which makes a positive contribution (to the sum over paths) then it is always possible to find another path infinitesimally close (on a classical scale) which makes an equal negative contribution. So, in the classical limit, the only paths that will contribute significantly (to the sum over paths) will be those for which a small change in path produces no change in S. This is true (to first order) for the paths which make S stationary (i.e. $\frac{\delta S}{\delta x} = 0$). These are just the classical paths. In the semi-classical approximation the path integral is proportional to $e^{\frac{i}{\hbar}S_{cl}}$ (where S_{cl} is the action evaluated on the classical path $x_{cl}(t)$). If there is more than one classical path then a sum of such terms is required.

2.5 Summary

The main points made in this chapter were:

1. Although Dirac's method was not really intended for the quantization of classical mechanical systems, it is generally assumed to be widely applicable. In fact it is not suitable for mechanical systems with nonholonomic constraints.

2. To quantize a constrained system using Dirac's procedure requires the successful completion of two stages: the first stage is to put the constrained system in Hamiltonian form, ready for quantization. The second stage is the quantization. For a mechanical system with nonholonomic constraints, even the results of the first stage are unsatisfactory. Quantization using Feynman's path integral seems much more promising. In the standard case this provides a direct route from the Lagrangian formulation of the

classical mechanics to the quantum mechanics. Another advantage is the intuitive picture provided by Feynman's formulation.

3. The most natural approach to path integral quantization (appendix C) is equivalent to the quantization of vakonomic mechanics. So the classical limit will not be correct when the constraints are nonholonomic (the vak motion will be obtained instead).

4. A quantum system is required which corresponds to a classical system with nonholonomic constraints. If the classical limit is to be correct then the standard approach is ruled out and some generalization of the path integral quantization must be found. There is no obvious direct approach to solving the problem, indeed, it is not clear if a solution exists. An investigation is required. It seems advisable to start with a simple system.

Chapter 3

Paraxial optics

3.1 Introduction

Faced with the problem of quantizing a system which lies outside the scope of standard quantization procedures, one approach is to consider what is meant by "quantization" and "taking the classical limit".

Quantum mechanics is a wave theory. The process of "taking the classical limit" is a reduction of a wave theory to a ray theory in the short wavelength limit [16]. In the case of quantization, the ray theory (classical mechanics) is known and it is required to find a wave theory consistent with this. It is well known that classical mechanics is analogous to ray optics and that quantum mechanics is analogous to wave optics. Consequently, quantization is analogous to the extension of geometrical optics to wave optics. If a specialisation is made to (2D) non-relativistic mechanics (where gradients of world-lines in space-time are assumed small) then one can go further. Paraxial optics (sometimes called Fresnel optics) and (2D) non-relativistic mechanics are mathematically identical provided identification is made between appropriate quantities in the two theories and the difference between the metrics of space (for optics) and space-time (for mechanics) is accounted for.

If, given a classical mechanical system, it is possible to construct a class of optical systems with corresponding ray dynamics, then applying wave optical methods to these physical systems is equivalent to quantization of the mechanical system [16].

More generally, visualisation of a system is often easiest when it is interpreted in terms of optics.

3.2 The limit

Paraxial optics is an approximation to optics for small gradients of rays in the same sense that non-relativistic mechanics is an approximation to relativistic mechanics for velocities $v \ll c$ (the speed of light in a vacuum). However, generally a slightly different view is taken: classical mechanics is a self-consistent description of mechanics and it is only necessary to work with relativistic mechanics if relativistic effects are likely to be important. The situation is exactly the same for paraxial optics.

An approximation is made in deriving paraxial optics from optics, but once the paraxial equations are obtained one is entitled to work entirely within the new self-consistent theory. For this reason the term paraxial "limit" will be preferred over "paraxial approximation". Just as the term "classical limit" is used in mechanics even though $\frac{v}{c}$ is finite.

3.3 Mechanics and optics

An instructive comparison between paraxial optics and non-relativistic mechanics is presented in [15]. It is worth summarising its contents here.

- Paraxial ray optics:

  a ray in a refracting medium of refractive index $n = \varphi + 1$ always makes a small angle with the z axis and obeys the paraxial ray equation

  $$\frac{d^2\underline{r}}{dz^2} = \underline{\nabla}_\perp \varphi \tag{3.1}$$

  where $\underline{\nabla}_\perp = \frac{\partial}{\partial \underline{r}}$

  or, equivalently, Fermat's principle of stationary optical length, $\delta S = 0$, where

  $$S[\underline{r}(z)] = \int_0^L \left[1 + \frac{1}{2}\left(\frac{d\underline{r}}{dz}\right)^2 + \varphi(\underline{r}(z), z)\right] dz \tag{3.2}$$

- Non-relativistic classical mechanics in 2D:

  the world-line of a non-relativistic particle moving in a plane under the influence of a potential V always makes a small angle with the ct axis and obeys Newton's law

  $$\frac{1}{c^2}\frac{d^2\underline{r}}{dt^2} = -\underline{\nabla}\left(\frac{V}{mc^2}\right) \tag{3.3}$$

  or, equivalently, the principle of stationary action, $\delta S = 0$, where

  $$S[\underline{r}(z)] = \int_0^{cT} \left[-1 + \frac{1}{2}\left(\frac{1}{c}\frac{d\underline{r}}{dt}\right)^2 - \frac{1}{mc^2}V(\underline{r}(t), t)\right] d(ct) \tag{3.4}$$

  Similarly,

- Paraxial wave optics:

  the wave field $\psi(\underline{r}, z)$ for a wave with frequency ck obeys the paraxial wave equation

$$\underline{\nabla}_{\perp}^2 \psi + 2k^2(1+\varphi)\psi = -2ik\frac{\partial \psi}{\partial z} \tag{3.5}$$

  $\psi(\underline{r}, z)$ is generated through Huygen's principle by the kernel

$$K(2,1) = \int\int_1^2 \exp\left(ik\int_0^L \left[1 + \frac{1}{2}\left(\frac{d\underline{r}}{dz}\right)^2 + \varphi(\underline{r}(z), z)\right] dz\right) d^{\infty}\underline{r}(z) \tag{3.6}$$

  should be compared with

- Non-relativistic quantum mechanics in 2D:

  the probability amplitude $\psi(\underline{r}, z)$ for a particle with "Compton wave-number" $\frac{mc}{\hbar}$ obeys the Schrödinger equation

$$\underline{\nabla}^2\psi + 2\left(\frac{mc}{\hbar}\right)^2\left(1 - \frac{V}{mc^2}\right)\psi = -2i\left(\frac{mc}{\hbar}\right)\frac{1}{c}\frac{\partial \psi}{\partial t} \tag{3.7}$$

  $\psi(\underline{r}, t)$ is generated through Huygen's principle by the kernel

$$K(2,1) = \int\int_1^2 \exp\left(-i\left(\frac{mc}{\hbar}\right)\int_0^{cT}\left[-1 + \frac{1}{2}\left(\frac{1}{c}\frac{d\underline{r}}{dt}\right)^2 - \frac{1}{mc^2}V(\underline{r}(t), t)\right] d(ct)\right) d^{\infty}\underline{r}(t) \tag{3.8}$$

Through out, it is assumed that, "back-tracking" paths are excluded from the path integrals (in both optical and mechanical cases). Also, the kernels obey the fundamental relations (with $1, 2, 3$ referring to an ordered sequence of events)

$$K(1', 1'') = \delta(1', 1'') \tag{3.9}$$

$$\int K^*(2, 1')K(2, 1'')d^2\underline{r}_2 = \delta(1'', 1') \tag{3.10}$$

$$\int K(3, 2)K(2, 1)d^2\underline{r}_2 = K(3, 1) \tag{3.11}$$

3.4 The wave equation

It is possible to obtain the paraxial wave equation for "free space" from the Helmholtz equation for the propagation of light in a vacuum in 3D space i.e.

$$\nabla^2\Psi + k^2\Psi = 0 \tag{3.12}$$

substituting

$$\Psi(x, y, z) = \psi(x, y, z)e^{ikz} \tag{3.13}$$

gives

$$\left[\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)\psi+2ik\frac{\partial\psi}{\partial z}+\frac{\partial^2\psi}{\partial z^2}\right]e^{ikz}=0 \tag{3.14}$$

for the paraxial case, one takes $k\frac{\partial\psi}{\partial z}\gg\frac{\partial^2\psi}{\partial z^2}$ so that the 3^{rd} term in equation (3.14) is negligible in comparison to the second, leading to the required (free space) paraxial wave equation

$$\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)\psi+2ik\frac{\partial\psi}{\partial z}=0 \tag{3.15}$$

3.5 Summary

The "physical" idea of quantization based on the analogy with optics is available even when standard prescriptions for quantization fail.

Chapter 4

Wave Equations

4.1 Introduction

The objective in this chapter is to investigate the possibility of obtaining a wave equation for nonholonomic systems. To aid visualisation of the system, paraxial optics is considered: instead of explicitly dealing with the non-relativistic mechanics of a particle in 2D (the idea is described in chapter 3). The question to be considered is whether it is possible to make progress towards a nonholonomic wave equation by considering the holonomic case, writing the result in a form independent of the surface and then making a formal generalisation to the nonholonomic case. The resulting equation can be "checked" by interpreting the terms present and seeking expected features.

In section 3.4, the paraxial wave equation for 3D "free space" was obtained. Now the corresponding equation for a 2D curved "constraint" surface is required. The first step is to write the Laplacian operator on this curved surface in terms of a differential operator which is as close as possible in direction to $\frac{\partial}{\partial z}$ but still "within" the surface. Once this has been done, the "paraxial substitution"

$$\Psi = \psi e^{ikz} \tag{4.1}$$

is made. The next stage is more complicated than in section 3.4 due to the presence of terms including $\frac{\partial}{\partial z}(\textit{spatial quantity})$ (e.g. $\frac{\partial x}{\partial z}$). These should be "small" since in the paraxial case the slope of paths is taken to be small. The "exploration" required to verify the most plausible "order of smallness" to associate with such terms is not included. Retaining terms to a consistent order gives the paraxial wave equation (the "Schrödinger equation" in the mechanical interpretation).

As a "check" on the result, the continuity equation corresponding to the "Schrödinger

equation" is obtained (in the standard way) and the terms are interpreted to check that they "make sense". One of the terms in the "Schrödinger equation" cancels out during the derivation of the continuity equation so a different method of "checking" this term is required. A simple special case is considered for this purpose. The WKB approximation scheme is used (there is only one space dimension within the "constraint" surface), in addition to taking the paraxial limit.

For the general case, the WKB approximation is applied to the "Schrödinger equation" and the resulting equations checked for plausibility by considering their form in special cases where the various terms are particularly simple.

Having investigated the holonomic case, a formal generalisation is made to the wave equation of the nonholonomic case. A simple special case (which will be considered again later) is used to test the plausibility of the result and an interpretation is attempted.

4.2 Investigation

In 3D space the equation governing the propagation of light in a vacuum is the Helmholtz equation

$$\nabla^2\Psi + k^2\Psi = 0 \tag{4.2}$$

The paraxial "limit" is obtained by substituting

$$\Psi(x,y,z) = \psi(x,y,z)e^{ikz} \tag{4.3}$$

with $k\frac{\partial\psi}{\partial z} \gg \frac{\partial^2\psi}{\partial z^2}$

With one holonomic constraint, light propagates in a 2D curved surface within the 3D Euclidean space. The Laplacian operator on a curved surface is given by

$$\frac{1}{\sqrt{g}}\frac{\partial}{\partial q^i}\left(\sqrt{g}\,g^{ij}\frac{\partial}{\partial q^j}\right)$$

where $g^{ij}g_{jk} = \delta^i_k \quad (i=1,2 \quad j=1,2 \quad k=1,2)$

and $g = \det g_{ij}$ (g_{ij} is the metric tensor)

To remove the explicit dependence on the coordinates (q^1, q^2), new differential operators (vector fields $\bar{a}, \bar{b}$) are defined. Requiring the two basis vectors to be orthonormal gives

$$\nabla^2\Psi = \bar{a}(\bar{a}\Psi) + \bar{b}(\bar{b}\Psi) + (-[\bar{a}\bar{b} - \bar{b}\bar{a}]_b)\bar{a}\Psi + ([\bar{a}\bar{b} - \bar{b}\bar{a}]_a)\bar{b}\Psi \tag{4.4}$$

where $[\bar{a}\bar{b} - \bar{b}\bar{a}] = [\bar{a}\bar{b} - \bar{b}\bar{a}]_a\bar{a} + [\bar{a}\bar{b} - \bar{b}\bar{a}]_b\bar{b}$ is the commutator of $\bar{a}$ and $\bar{b}$

If it is further required that one of them is in a transverse direction, tangent to curves of

constant z on the surface, then their form is determined uniquely in terms of the coordinates on the surface, one of which is chosen to be the z coordinate of the 3D Euclidean space. These specific operators are

$$\bar{u} = \frac{1}{\sqrt{g_{\alpha\alpha}}}\frac{\partial}{\partial\alpha}\Big)_z \tag{4.5}$$

$$\bar{v} = \sqrt{\frac{g_{\alpha\alpha}}{g}}\frac{\partial}{\partial z}\Big)_\alpha - \frac{g_{\alpha z}}{\sqrt{g_{\alpha\alpha}g}}\frac{\partial}{\partial\alpha}\Big)_z \tag{4.6}$$

where $g_{\alpha\alpha}, g_{\alpha z}$ are components of the 2×2 metric tensor describing the surface $x = x(\alpha, z)$, $y = y(\alpha, z)$, z and its coordinate system.

Working in the paraxial regime and neglecting terms small compared to $k\frac{\partial}{\partial z}(spatial\ quantity)$ gives

$$2ik\frac{1}{\bar{v}z}(\bar{v}\psi) + \bar{u}(\bar{u}\psi) + ik\frac{1}{\bar{v}z}[\bar{u}\bar{v} - \bar{v}\bar{u}]_u\psi + k^2\frac{1}{(\bar{v}z)^2}(1 - (\bar{v}z)^2)\psi = 0 \tag{4.7}$$

which is analogous to a time dependent Schrödinger equation. The corresponding continuity equation is

$$\frac{d}{dt}\int \psi^*\psi\sqrt{g_{\alpha\alpha}}d\alpha + \int\left(\bar{u}\left(-\frac{g_{\alpha z}}{\sqrt{g_{\alpha\alpha}}}\psi^*\psi + \frac{1}{2ik}[\psi^*(\bar{u}\psi) - (\bar{u}\psi^*)\psi]\right)\right)\sqrt{g_{\alpha\alpha}}d\alpha = 0 \tag{4.8}$$

This is analogous to the standard result except for the second term which accounts for a bulk flow of probability if the $z = constant$ and $\alpha = constant$ directions are not orthogonal. This result is independent of the form of the last term in the wave equation, and therefore does not provide a "check" for this term in the "Schrödinger equation". However, this term may be checked for consistency by considering the classical limit. We know what to expect for the "path length" and substituting this into the equation allows a consistency check to be made. If a simple special case is considered then one can rederive the equations with the benefit of clear geometrical guidance. If the relevant special case of the general result agrees with this, then confidence in the general result is increased. Choosing an inclined plane with normal vector in the $(0, 1, -y_z)$ direction (y_z is the derivative of y with respect to z), returning to the Helmholtz equation, making the substitution $\Psi = Ae^{ikQ}$ and retaining only the highest order terms in k gives

$$(\bar{v}Q)^2 + (\bar{u}Q)^2 = 1 \tag{4.9}$$

then substituting $Q = z + q$ and neglecting the $(\bar{v}q)^2$ term since it is small compared to $\bar{v}q$ gives

$$2(\bar{v}q) + (\bar{u}q)^2 = y_z^2 \tag{4.10}$$

when higher order terms in $\frac{\partial}{\partial z}(spatial\ quantity)$ are neglected. Since the expression for the "path length" q is known, it is possible to verify that this expression is consistent, to within the approximation used, by substituting for q (on the left side) and comparing the result with y_z^2. A similar procedure may be followed for the general case by substituting $\psi = Ae^{ikq}$. Considering the real part of the resulting equation, retaining only the highest order terms in k and dividing through by A gives

$$\frac{2}{\bar{v}z}(\bar{v}q) + (\bar{u}q)^2 = \frac{1-(\bar{v}z)^2}{(\bar{v}z)^2} \tag{4.11}$$

which is the general case corresponding to equation (4.10). The imaginary part gives

$$\frac{1}{\bar{v}z}\bar{v}A + (\bar{u}q)(\bar{u}A) + \frac{1}{2}\bar{u}(\bar{u}q)A + \frac{1}{2(\bar{v}z)}[\bar{u}\bar{v} - \bar{v}\bar{u}]_u A = 0 \tag{4.12}$$

In the holonomic case the first two terms combine to give $\frac{dA}{dz}$. The third term is associated with the divergence of rays within the surface , it is zero when the surface is a cone. The rays are then along generators of the cone. The fourth term is non-zero for the cone but zero for a flat surface so it may be interpreted as a measure of the "divergence" of the surface itself.

In the holonomic case the operators $\bar{u}$ and $\bar{v}$ are defined in terms of the surface and coordinate system. To extend the wave equation to the nonholonomic case, a more general interpretation of these operators is required. Again taking $\bar{u}$ to have no component in the z direction and to be normalized, the additional requirement that its component vector be orthogonal to the local value of the normal vector, $\underline{n}$, specifying the planelet field (i.e. "within the planelet") gives

$$\bar{u} = \frac{n_y}{\tau}\frac{\partial}{\partial x} - \frac{n_x}{\tau}\frac{\partial}{\partial y} \tag{4.13}$$

where $\tau = \sqrt{n_x^2 + n_y^2}$

Taking $\bar{v}$ to be normalized and orthogonal to $\bar{u}$ and its component vector to be orthogonal to $\underline{n}$, gives

$$\bar{v} = \frac{n_x n_z}{n\tau}\frac{\partial}{\partial x} + \frac{n_x n_z}{n\tau}\frac{\partial}{\partial y} + \frac{\tau}{n}\frac{\partial}{\partial z} \tag{4.14}$$

where $n = \sqrt{n_x^2 + n_y^2 + n_z^2}$

The wave equation resulting from using these definitions may be checked using a simple nonholonomic special case.

For the case $\underline{n} = \underline{n}(z)$ (only) and $n_z = 0$ the planelets join to form strips of infinite length but of infinitesimal width in the z direction. These stacks of strips rotate as z increases. The last term in the wave equation vanishes. The commutator's component

vector is in the $\underline{n}$ direction which contrasts with the holonomic case where the commutator was always "within the surface". So the u component which occurs in the wave equation is zero. Consequently the wave equation takes the simple form.

$$2ik\frac{\partial\psi}{\partial z} + \bar{u}(\bar{u}\psi) = 0 \tag{4.15}$$

where $\bar{u}$ is like a partial derivative with respect to distance along a "planelet strip". This is analogous to a Schrödinger equation with a constant potential but in a rotating coordinate frame. The problem is that a nonholonomic constraint should not reduce the dimension of the accessible position space, so $\psi = \psi(x, y, z)$, but the wave equation does not seem to specify what ψ does in the $\underline{n}$ direction. This is perhaps not surprising when the method of derivation is considered. The original surface is just one of a "stack" of surfaces filling 3D space. However simply introducing a variable labelling the surfaces within the stack will not solve the problem.

4.3 Summary

Although a definitive nonholonomic wave equation has not been obtained, insight has been gained into some potential difficulties in achieving this goal.

The approach used here is novel, but the problem of obtaining the Schrödinger equation for a particle on a surface or a curve appears in the literature, including questions of the dimensionality of the wavefunction. The fundamental distinction is between applying the constraints before quantization or after quantization. If the constraints are applied to the classical system (before quantization) then the dimension of the wavefunction obtained after quantization is reduced. If the constraint is applied by imposing a deep potential well on the quantum system (i.e. after quantization), then the wavefunction depends on the full space but the part depending on coordinates within the surface is often treated separately [19, 24].

The approach presented in this chapter fits into such a classification in so much as the dimension of the wavefunction is reduced when a holonomic constraint is present. The "confining potential well" approach is often advocated for systems where the constraint arises from a real physical system, rather than a mathematical formalism. Since the problem considered here is based on real (although idealised) mechanical systems, it might seem that a development of the "confining potential well" approach might be more appropriate. However, the fact that nonholonomic constraints are necessarily velocity dependent rules out a direct extension of such a procedure.

Chapter 5

A simple non-holonomic system

5.1 Introduction

Faced with a potentially difficult investigation, it is advisable to start with a simple system. If progress is made with this, then a generalisation can be attempted. This is likely to be easier than beginning with the most general case. This chapter introduces a simple system. A study of this system is presented in subsequent chapters.

5.2 The system

The system is a point mass in a (horizontal) plane x, y which is subject to a single non-holonomic constraint $f(x, y, \dot{x}, \dot{y}, t) = 0$ of the simple form $n_x(t)\dot{x} + n_y(t)\dot{y} = 0$ (easily generalizable to $n_x(t)\dot{x} + n_y(t)\dot{y} + n_t(t) = 0$). A physical realization of this is the variable radius rolling disc described earlier (in section 1.3). The constraint being independent of x, y and linearly dependent on $\dot{x}, \dot{y}$ means that the classical mechanics, both vakonomic and ordinary nonholonomic, can be solved explicitly, as can the quantum mechanics in the standard path integral (ie vakonomic quantum) formulation. This allows attention to be focused on attempting to find a formulation for quantum ordinary mechanics.

For this special case the planelets join up to form strips of infinitesimal length in the time direction. Also, the results in this chapter take $\underline{n}$ to be normalized, i.e. $n_x^2 + n_y^2 = 1$

For this simple system, the space-time version of the "holonomicity condition" (1.5) reduces to

$$n_x \dot{n}_y - n_y \dot{n}_x = 0 \tag{5.1}$$

but $\underline{n}$ may be written in the form

$$(n_x, n_y) = (-\sin\phi(t), \cos\phi(t)) \tag{5.2}$$

where ϕ is some (unspecified in the general case) function of time.

So that

$$(\dot{n}_x, \dot{n}_y) = (-\dot{\phi}\cos\phi(t), -\dot{\phi}\sin\phi) \tag{5.3}$$

and

$$n_x\dot{n}_y - n_y\dot{n}_x = \dot{\phi} \tag{5.4}$$

In other words, the constraint is only holonomic if $n_x = constant$ and $n_y = constant$. When time dependence is present (i.e. $\phi \neq constant$) the constraint is nonholonomic.

5.3 The classical mechanics

5.3.1 Vakonomic solution

The special case for $n_t = 0$ and $\underline{n}$ normalized, of the result in appendix D is

$$\Delta\underline{r}(t) = \left[\int_{s=0}^{t}(1 - \underline{n}\,\underline{n})ds\right]\left[\int_{u=0}^{T}(1 - \underline{n}\,\underline{n})du\right]^{-1}\Delta\underline{r}_T \tag{5.5}$$

where $\Delta\underline{r}_T = \underline{r}(T) - \underline{r}(0)$ is the required displacement , 1 is the unit matrix and $\underline{n}\,\underline{n}$ is the outer product of two $\underline{n}$ vectors.

If the final displacement is specified (i.e. $\Delta\underline{r}_T$) then the vak solution tells us "how to get there" (i.e. $\Delta\underline{r}(t)$).

5.3.2 Nonholonomic solution

The nonholonomic equations of motion are $\ddot{\underline{r}} = -(\dot{\underline{n}}.\dot{\underline{r}})\underline{n}$. The constraint $\underline{n}.\dot{\underline{r}} = 0$ ensures that the velocity is $v(n_y, -n_x)$. Conservation of energy $\dot{\underline{r}}.\ddot{\underline{r}} = 0$ means that its magnitude $|v|$ is constant. Integration of this velocity gives

$$\Delta\underline{r}(t) = (\underline{n}(0) \wedge \dot{\underline{r}}(0))\int_0^t (n_y(s), -n_x(s))ds \tag{5.6}$$

where the factor outside the integral is the signed initial velocity.

Considering a particular initial point but all possible values of v gives a "classical fan" of rays (section 1.5.5). At any given time the classical fan defines a straight line which is not in general parallel to the "planelet strips". It is hoped that this "classical fan" will become an important feature of a correct quantum solution as the classical limit is taken.

5.4 The quantum mechanics

5.4.1 The vakonomic propagator

The path integral may be evaluated directly or the classical vakonomic solution may be used to obtain the (vak) "classical" action. The special case for $n_t = 0$ and $\underline{n}$ normalised, of the result in appendix D is

$$\begin{aligned} K &= \int \int_{\underline{r}(0),0}^{\underline{r}(t),t} e^{\frac{i}{\hbar}\int_{\tau=0}^{t}(\frac{m}{2}\dot{\underline{r}}^2+\lambda(\underline{n}.\dot{\underline{r}}))d\tau} d^{\infty}\underline{r}(\tau) d^{\infty}\lambda(\tau) \\ &= \frac{2\pi m}{i\hbar\sqrt{\det(M)}} e^{\frac{im}{2\hbar}\Delta\underline{r}M^{-1}\Delta\underline{r}} \end{aligned} \tag{5.7}$$

where M is the matrix $[\int_{\tau=0}^{t}(1-\underline{n}\,\underline{n})d\tau]$

The vakonomic propagator is not expected to give the correct classical limit but it is worth considering how it does in fact behave. The classical limit ("$\hbar \to 0$") means that $\hbar$ is small compared to a typical classical action. In this limit equation (5.7) shows that K oscillates rapidly as $\Delta\underline{r}$ is changed. This might be acceptable behaviour for a "nonholonomic propagator" if the oscillation were least rapid on the classical fan. If the *phase = constant* curves were ellipses which became increasingly elongated around the line of the classical fan as "$\hbar \to 0$", for example. The condition $\Delta\underline{r}M^{-1}\Delta\underline{r} = constant$ does define a set of ellipses, so the required anisotropy is present. However, finding the eigenvectors of M^{-1} shows that their axes do not lie along the line of the classical fan. The vakonomic propagator was not expected to give the correct classical limit, but this example illustrates the type of test any proposed nonholonomic propagator must pass to be acceptable.

Having evaluated the path integral for $K(\underline{r}_2, t_2; \underline{r}_1, t_1)$ it is straightforward to obtain the corresponding differential equation [10, 32]

$$i\hbar\frac{\partial\psi}{\partial t} = -\frac{\hbar^2}{2m}\left(n_y^2\frac{\partial^2}{\partial x^2} - 2n_x n_y\frac{\partial^2}{\partial x\partial y} + n_x^2\frac{\partial^2}{\partial y^2}\right)\psi \tag{5.8}$$

Since $\underline{n}$ is normalised and depends only on time, this agrees with the result (4.15) with $k = \frac{mc}{\hbar}$. So in this special case, the wave equation obtained by generalizing the holonomic result agrees with that derived from the standard path integral (i.e. the vakonomic result). This wave equation may represent "quantum mechanics within the planelet strip".

5.4.2 The nonholonomic propagator

The natural approach to incorporating constraints into the standard path integral gives "vakonomic mechanics" so a new approach is required. The goal is some form of "general-

ized path integral", for the simple nonholonomic system, which gives the correct classical mechanics in the classical limit. An approach to this problem using a model system is described in the next chapter.

5.5 Summary

The purpose of this chapter is to introduce the simple nonholonomic system which is considered in the remaining chapters. Also included are the results which may be obtained for this simple system, and some discussion of these. The results given are the classical and quantum vakonomic solutions and the classical (correct) nonholonomic solution. The "set" is not complete because no result is given for the "quantum nonholonomic" solution. The search for such a solution is the subject of the following chapters.

The results presented in this chapter are not likely to occur in the literature. This is partly due to the idealised nature of the system considered: its simplicity provides the opportunity to find explicit solutions. Also, the "quantum vakonomic" result presented here is equivalent to *solution* of the differential equation which would result from the standard "Dirac procedure". Often the goal is simply to obtain the differential equation.

Chapter 6

The model

6.1 Introduction

The quantization of a system subject to a nonholonomic constraint using the standard Feynman path integral formulation gives "quantum vakonomic mechanics". The correct classical mechanics is not obtained in the classical limit. The goal is to modify the standard path integral formulation so that the constraints are incorporated and the classical limit is correct. The idea is to include the constraints at the most fundamental level in the construction of the path integral.

The problem of constructing the propagator for a finite time interval is broken into two "sub-problems". The first sub-problem is to find a suitable quantity to represent the propagator for an infinitesimal time interval. The second sub-problem is to "compose" these quantities into an expression for the propagator for a finite time interval. This is just the same as in the standard case (section 2.4). A "suitable quantity" is an approximation to the propagator for a small time interval, ϵ, valid to first order in ϵ. This follows because the expression for the propagator for (finite) time Δt (i.e. equation (2.7) if $\Delta t = t_b - t_a$) effectively contains $\frac{\Delta t}{\epsilon}$ factors of this type. If an error of order ϵ^2 is made in each, the resulting error will not accumulate beyond the order $\epsilon^2 \left(\frac{\Delta t}{\epsilon}\right)$ (i.e. $\epsilon \Delta t$) which vanishes in the relevant limit (i.e. $\epsilon \to 0$).

The proposed approach is to apply the constraint directly to the propagator for an infinitesimal time interval. Of course, if "composition" to form the finite time interval propagator is to be achieved then one must consider errors to check that the cumulative error will not be too large. However, in the present chapter, the focus is on the first sub-problem — to find a plausible way to incorporate the constraint into the infinitesimal

propagator.

6.2 Introducing the constraint

The constrained system to be considered is the "simple nonholonomic system" introduced in chapter 5. The "picture" of the constraints appropriate to this case was mentioned in section 5.2. Now attention is to be focused on an infinitesimal time interval. Over an infinitesimal time interval, the "constraint normal vector" $\underline{n}$ may be taken to be constant. This vector defines a direction normal to each "planelet" (i.e. "$(\underline{n},0)$" in 3D space-time, where "$(\underline{n},0)$" is defined to mean $(n_x, n_y, 0)$). For the simple system, it is independent of position, so the planelets all have the same orientation and "join up" to form strips of infinitesimal length in the time direction of 3D space-time. So for this case of an infinitesimal time interval, the constraints may be represented by a stack of strips (parallel to each other) infinitely densely packed within this section of space-time. In fact, to follow the standard path integral construction, it is necessary to "go back a step" and consider a small time interval ϵ, the infinitesimal case can be obtained later by taking the limit $\epsilon \to 0$.

With the "infinitesimal" time interval extended to length $\epsilon \neq 0$ the stack of "constraint strips" is no longer infinitely densely packed — there is a small separation, a, between each adjacent pair of planes in the stack. This means that when the limit $\epsilon \to 0$ is taken, there is a choice to be made between:

1. $\frac{a}{\epsilon} \to 0$ as $\epsilon \to 0$
2. $\frac{a}{\epsilon} \to \infty$ as $\epsilon \to 0$
3. $a \sim \epsilon$ as $\epsilon \to 0$

any of these could be chosen provided that the requirement that $a \to 0$ as $\epsilon \to 0$ is met (so that the stack of strips is infinitely dense in the limit). It is possible that this apparent extra "freedom" in fact has no effect on the limiting case. It is also possible that only one case can give the correct behaviour and that another case is associated with vakonomic type behaviour.

There remains the task of proposing a "mechanism" by which the constraints are enforced. The constraint is that there should be no motion in a direction perpendicular to the "constraint (planelet) strips". This need only be rigidly enforced in the limit (i.e. $\epsilon \to 0$). A way to realise the constraint is to treat the constraint strips as rigid barriers. Moving to the terminology of optics as described in chapter 3, one might consider that the rays are

guided between pairs of strips acting as "wave-guides". In this case it is clear how waves will behave — so a possible "quantization" has been found (chapter 3 and [16]). So, in this scheme, the constraints are incorporated as a requirement that the wave-function must be zero on the "constraint strips" (described as "mirror plane strips" in the optical terminology). Between these strips are lanes of "free space". The stacks of constraint strips for two separate times are illustrated in figure 6.1, although the strips are of infinite length, they have been truncated in the diagram. The spacing of the strips within the stack, a, is not a function of time.

Before returning (in section 6.3) to the first sub-problem, it is worth considering the second sub-problem, i.e. the question of "composition". For the next time interval ("stage") the entire stack of strips ("Venetian blind") is rotated slightly (by an angle $\Delta\theta$, with $\Delta\theta \sim \epsilon$ as $\epsilon \rightarrow 0$ since the rate of rotation is finite). At the interface between stages (i.e. a $t = constant$ "interface plane" — examples are included in figure 6.1) the "leading edges" of the first set of constraint strips intersect with the "trailing edges" of the second set to form rhombus shaped (2D) "boxes" or "unit cells" (on the interface plane). A Feynman path may be labelled by a set of numbers: its coordinates at each "time-slice". Since the presence of constraints yields a grid of rhombuses, it is natural to use a list of the rhombuses through which a path passes to provide such a specification.

6.3 Single stage propagation

The single stage propagator will be obtained for a time interval (ϵ) of general length. Eventually, the limit $\epsilon \rightarrow 0$ will be taken (as in the standard case), since a single stage is of infinitesimal duration. An advantage of working with finite ϵ is that one can use all the techniques of path integration to aid derivation of the desired formula.

During a stage the particle is confined between a pair of "planes" (the constraint strips) which have $(\underline{n}, 0)$ as their normal vector. It is in a potential which is translationally invariant in the direction perpendicular to $\underline{n}$ and gives an infinite square well (with wall separation a) in the $\underline{n}$ direction.

The (2D) propagator for a single "stage" lasting for a time interval ϵ may be factorised into the 1D propagator for motion parallel to $\underline{n}$ and the 1D propagator for motion in the (orthogonal) direction between the planes (i.e. free space propagation). For the propagator for motion parallel to $\underline{n}$ (with initial and final points between the same pair of planes) the path integral reduces to a sum over straight classical paths for a particle bouncing between

the walls, since there is free space between the walls. The (type of) path segments to be summed over are illustrated in figure 6.2. If the initial and final points are separated by a "constraint plane" then the propagator is zero.

When the initial and final points are in the same "lane" (i.e. there is no "constraint plane" between them), an alternative way to obtain the expression for the 1D propagator (for motion parallel to $\underline{n}$) is to remove the constraining "planes" (boundaries) of the "lane" in which the particle is moving and add "images". Specifically, a charged particle between parallel conducting planes may be considered: the planes are replaced by image charges, and a contribution from each image charge included to give the 1D single stage propagator

$$K(\Delta x, \beta) = \sqrt{\frac{\nu}{\pi i}} \left(\sum_{n=-\infty}^{\infty} e^{i\nu(\Delta x + 2an)^2} - \sum_{N=-\infty}^{\infty} e^{i\nu(\Delta x + 2\beta + 2aN)^2} \right) \tag{6.1}$$

where $\Delta x = x_{final} - x_{initial}$

$$\nu = \frac{m}{2\hbar\epsilon} \qquad \text{(mechanics)} \tag{6.2}$$

or

$$\nu = \frac{k}{2c\epsilon} \qquad \text{(optics)} \tag{6.3}$$

ϵ = duration of stage in time

m = mass of particle

$\beta = x_{initial}$ if the initial and final points within the same "lane", otherwise $K = 0$.

Figure 6.3 shows how the first few image charges are constructed from a positive charge in the central "lane", the pattern in fact extends to infinity in both directions — in the same way that an object placed between parallel mirrors has an infinite number of images in both directions. The horizontal lines in the diagram are for construction purposes, they can be ignored as far as the "paths of the image charges" are concerned. Applying elementary geometry to the diagram is sufficient to determine the displacements, which are required for the "free space" $\exp\left(i \times constant \times \frac{(displacement)^2}{\Delta t}\right)$ terms in equation (6.1).

The sums in equation (6.1) do not give a finite result. Generally, this is true even when they are combined. Truncated versions are related to curlicues [6] and thus have complicated parameter dependence. Some insight can be gained by considering the Wigner function for the 1D wavefunction composed of positive and negative "delta combs". Free space is a case when its time evolution is given simply by a shear of phase space.

6.4 Summary

This chapter has initiated the investigation of the simple system introduced in chapter 5. In particular, a "physical" realisation of the constraints has been suggested, which is consistent with the geometrical picture of this simple system. The advantage of basing a model on a physical system (a charged particle between parallel conducting planes in this case) is that it can be investigated using standard techniques from physics. However, it is not guaranteed that the model will exhibit the behaviour desired for a nonholonomic system. If investigation of the model suggests that its behaviour is unsuitable, then it is necessary to modify the model or possibly to reject it altogether. This is just standard modelling procedure. In this work it is also advisable to consider whether the correct question is being asked of the model.

From the investigation of a single stage in this chapter, it seems that a re-assessment is indeed required.

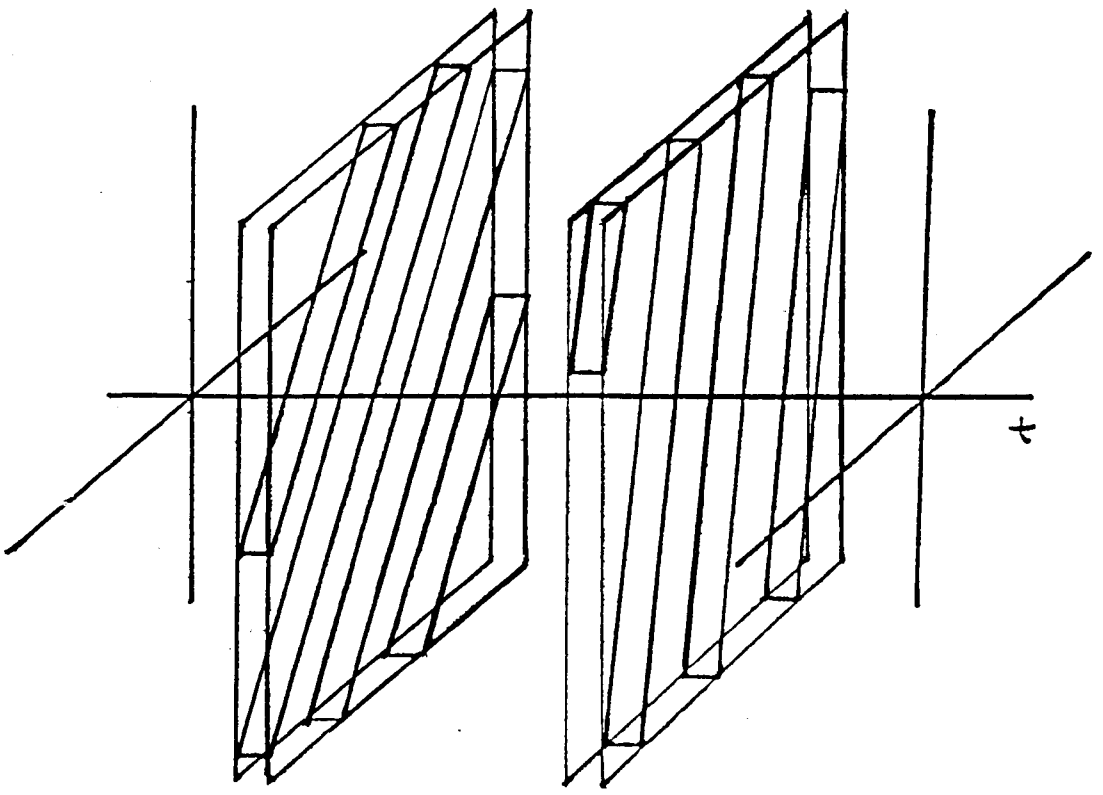


Figure 4.1: Two "infinitesimal" stages with "interface planes"

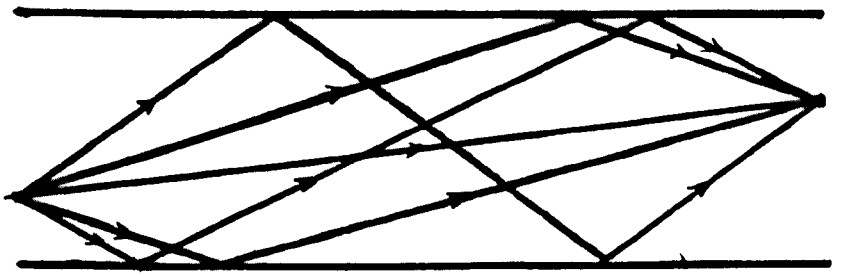


Figure 4.2: Classical bouncing paths

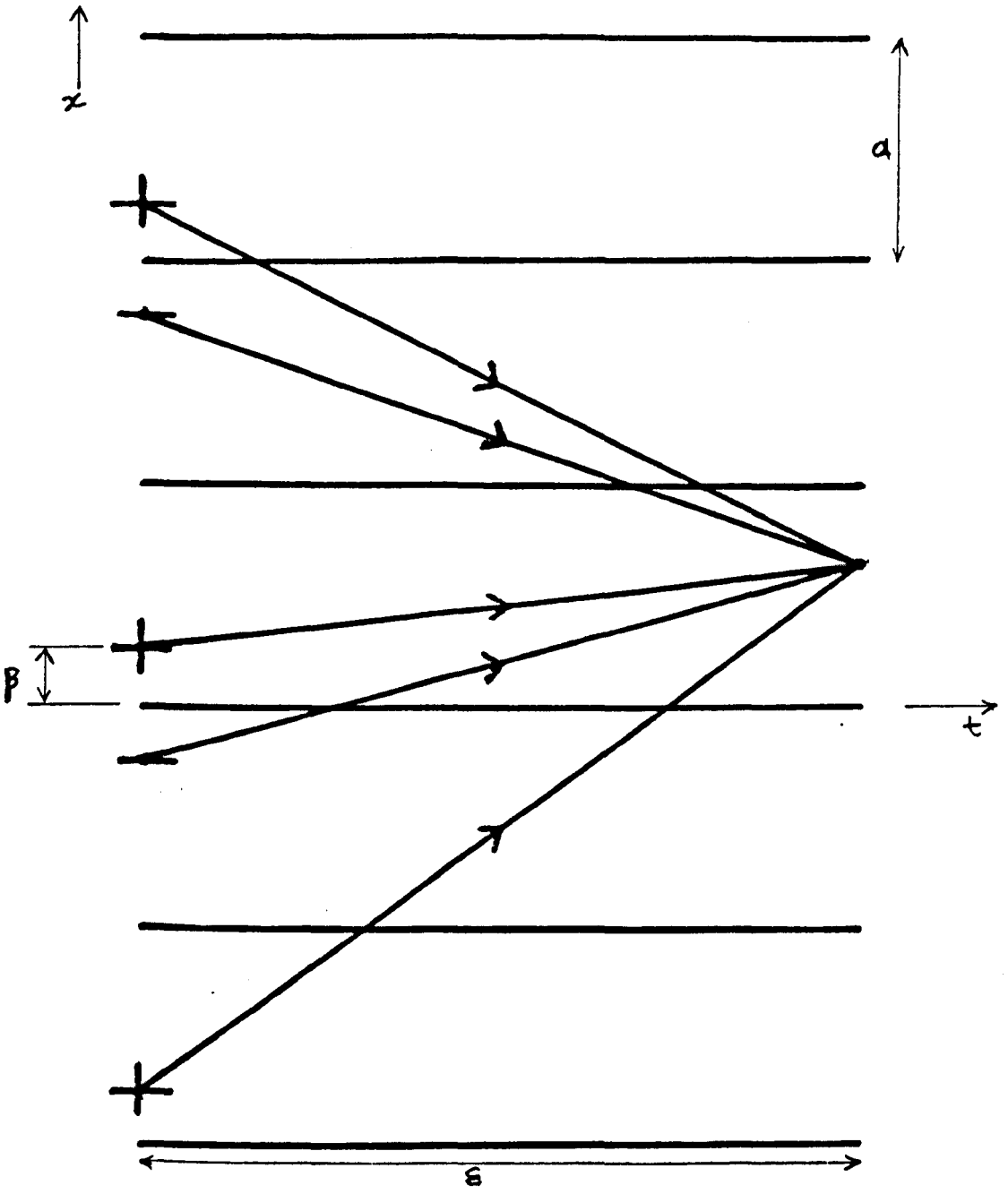


Figure 4.3: The image charges construction (5 "lanes" only)

Chapter 7

Modes

7.1 Introduction

This chapter introduces a way of performing calculations based on the model in chapter 6. The focus of the calculations is different but the model is the same. The "singular" nature of the propagator (6.1) suggests considering a transition amplitude [10] instead. Specifically, problems resulting from propagating from a specified point to another point should be avoided by considering instead the transition from a specified mode to another mode. During a stage the constraint is realised by a wave-guide, so mode analysis can be applied (the form of the propagator in the modes scheme is given in appendix F). Figure 7.1 gives a schematic representation of some modes on a section through the stack of constraint strips. At the intersections between stages the wavefunction overlaps are considered on the rhombus "unit cells". Since "composition" is considered in more detail in this chapter, the full 2D formulae will be given: with the ("free space") direction parallel to the "constraint strips" included.

7.2 The transition amplitude

Whereas the propagator provides an expression to go "from point to point", the transition amplitude is "from mode to mode". The process:

- start with a given mode at an interface between stages
- calculate the overlap with the modes in the next stage
- propagate a distance $\Delta z (> 0)$ in the direction of the z-axis

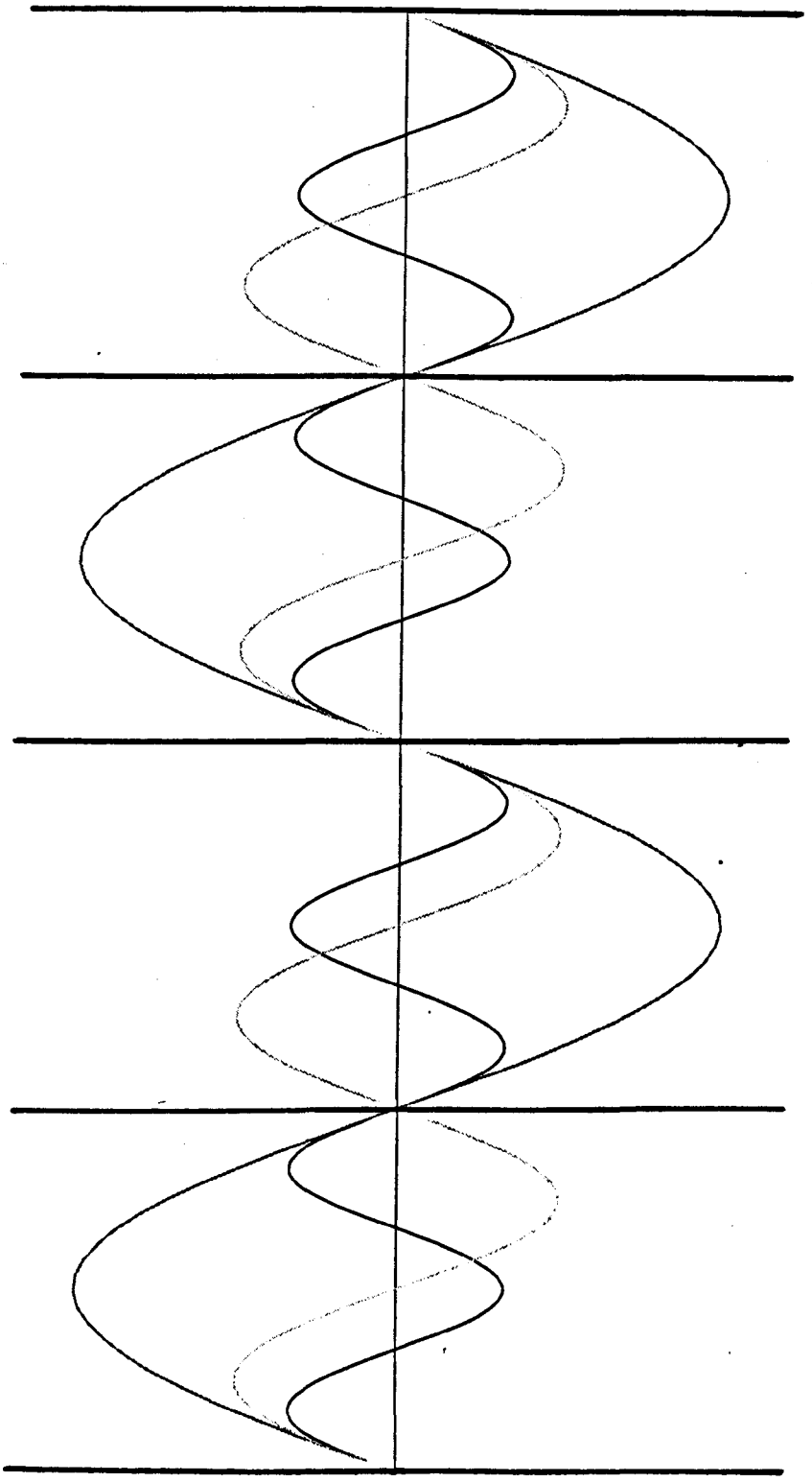


Figure 5.1: Lowest 3 modes in x-t plane

- at the interface calculate the overlap with a specified mode in the next stage

is represented by the transition amplitude

$$\begin{aligned} T(n_3, k_{y3}; n_1, k_{y_1}) &= \langle \chi(n_3, k_{y3}) | \psi(n_1, k_{y_1}) \rangle \\ &= \int\int \chi^*(\underline{r}''; n_3, k_{y3}) K(\underline{r}'', z''; \underline{r}', z') \psi(\underline{r}'; n_1, k_{y_1}) d^2\underline{r}' d^2\underline{r}'' \end{aligned} \quad (7.1)$$

where

$$K(\underline{r}'', z''; \underline{r}', z') = \sum_{n_2=1}^{\infty} \int_{-\infty}^{\infty} \phi(\underline{r}''; n_2, k_{y2}) \phi^*(\underline{r}'; n_2, k_{y2}) e^{ik_{z_2}\Delta z} dk_{y2}$$

for $z'' - z' = \Delta z > 0$

$$k_{z_i} = k\left(1 - \frac{(k_{x_i}^2 + k_{y_i}^2)}{2k^2}\right) \quad (7.2)$$

in the paraxial case considered here

and

$$\psi(\underline{r}; n_i, k_{y_i}) = \left(\sqrt{\frac{2}{a}} \cos \frac{n_i \pi}{a}(x_i - x_{ci})\right)\left(\frac{1}{\sqrt{2\pi}} e^{ik_{y_i} y_i}\right) \quad (7.3)$$

(with similar expressions for for χ and ϕ)

x_{ci} is the value of the x_i coordinate for the centre of a lane (of width a)

The x_i coordinate axes are in the direction of the normal vector of stage "i" and the y_i coordinate axes are perpendicular to the corresponding x_i axes (i.e. parallel to the constraint planes of stage "i"). Consequently, the coordinate system of each stage is rotated by an angle $\Delta\theta$ relative to that of the previous stage. The number of dashes label the interface at which the overlap is evaluated. Upon adding an extra propagation stage and interface the transition amplitude becomes

$$\begin{aligned} &T(n_4, k_{y4}; n_1, k_{y_1}) \\ &= \int\int\int \chi^*(\underline{r}'''; n_4, k_{y4}) K(\underline{r}''', z'''; \underline{r}'', z'') K(\underline{r}'', z''; \underline{r}', z') \psi(\underline{r}'; n_1, k_{y_1}) d^2\underline{r}' d^2\underline{r}'' d^2\underline{r}''' \end{aligned} \quad (7.4)$$

For a continuously varying normal vector, both Δz and $\Delta\theta$ are infinitesimal, so this procedure must be repeated an infinite number of times to achieve a finite displacement in the z direction. By definition, if the angular frequency of rotation of the stack of strips ($\dot{\theta}$) is finite then $\Delta\theta \sim \dot{\theta}\Delta z$ as $\Delta z \to 0$. However, there is also the limiting behaviour of a to be considered. If $a \sim \Delta\theta^p$ (where $p > 0$) then the freedom to choose p allows a range of systems to be generated, ideally including models for both quantum ordinary nonholonomic and vakonomic cases (section 6.2). Since the wavefunction is zero on the

planelet strips it is necessary to divide by a power of a to prevent a zero result in the limit $a \to 0$. Otherwise this limit might suggest that the wavefunction is zero everywhere.

There is no reason why a particular position in space (configuration) of the stacks of strips should be preferred, so an average should be taken over shifts of each stack in the direction of its normal vector. Only shifts in the range 0 to a need be considered since each stack is periodic with period a under such shifts. If the lanes are labelled by integer variables (j_i) then a particular rhombus is specified by the values of j in the stages on either side of the interface. The integral over the whole interface plane may be broken down into an integral over a single general rhombus and a sum over all rhombuses. The position in space of a given rhombus is changed by a shift but if the formula for the overlap at an interface is written in terms of the lane labels then it is shift independent. Specifically, integrating over a rhombus unit cell at the first interface gives

$$\begin{aligned} & I(n_2, k_{y_2}, j_2; n_1, k_{y_1}, j_1) \\ &= \int_{\Delta x_1'=-\frac{a}{2}}^{\frac{a}{2}} \int_{y_1'=L_-(\Delta x_1')}^{L_+(\Delta x_1')} \cos\left(\frac{n_2\pi}{a}\Delta x_2'\right) \cos\left(\frac{n_1\pi}{a}\Delta x_1'\right) \exp(i(k_{y_2}y_2' - k_{y_1}y_1'))dy_1'd(\Delta x_1') \\ &= \frac{a^2}{\sin\Delta\theta} e^{i(j_1aA + j_2aB)} \frac{4\pi^2 n_1 n_2}{[(Aa)^2 - (n_1\pi)^2][(Ba)^2 - (n_2\pi)^2]} \cos\frac{Aa}{2} \cos\frac{Ba}{2} (-1)^{\frac{1}{2}(n_1+n_2-2)} \end{aligned} \tag{7.5}$$

for n_1, n_2 odd

where

$L_\pm(\Delta x_1') = y_{c1}' \pm \frac{a}{2}\csc\Delta\theta - \Delta x_1' \cot\Delta\theta$

$\Delta x_i' = x_i' - x_{ci}'$

$x_{ci}' = j_i a$

$A = k_{y_1}\cot\Delta\theta - k_{y_2}\csc\Delta\theta$

$B = k_{y_2}\cot\Delta\theta - k_{y_1}\csc\Delta\theta$

In the derivation of this result (appendix E) unnecessary complication is avoided by taking shifts to be zero from the beginning.

7.3 Composition of stages

The combination of a rhombus overlap integral at an interface followed by a single stage of propagation within a lane is, for example

$$J(n_2, k_{y_2}, j_2; n_1, k_{y_1}, j_1) = I(n_2, k_{y_2}, j_2; n_1, k_{y_1}, j_1) \exp\left(ik\Delta z - \frac{i\Delta z}{2k}\left[\left(\frac{n_2\pi}{a}\right)^2 + k_{y_2}^2\right]\right) \tag{7.6}$$

Two such steps are composed using

$$J(n_3,k_{y_3},j_3;n_1,k_{y_1},j_1)$$
$$= \sum_{j_2=-\infty}^{\infty}\sum_{n_2=-\infty}^{\infty}\int_{k_{y_2}=-\infty}^{\infty} J(n_3,k_{y_3},j_3;n_2,k_{y_2},j_2)J(n_2,k_{y_2},j_2;n_1,k_{y_1},j_1)dk_{y_2} \tag{7.7}$$

In order to compare this with the original transition amplitude it is necessary to note that the j_i variables were introduced to label rhombuses at interfaces, so each is really two separate labels. When both are summed over, as they are in the middle of a multiple stage expression, they may be combined but at the ends one member of the pair is missing. If it is taken as implicit that the meaning of a sum over j_i is slightly different for the "end sums" then the comparison may be written

$$T(n_3,k_{y_3};n_1,k_{y_1})\exp\left(ik\Delta z - \frac{i\Delta z}{2k}\left[\left(\frac{n_2\pi}{a}\right)^2 + k_{y_2}^2\right]\right)$$
$$= (\pi a)^{-3}\sum_{j_1=-\infty}^{\infty}\sum_{j_3=-\infty}^{\infty} J(n_3,k_{y_3},j_3;n_1,k_{y_1},j_1) \tag{7.8}$$

So an approach to composing steps based on evaluating the rhombus overlap integrals first allows some progress to be made but leaves summation over n_i and integration over k_{y_i} still to be done. It is difficult to evaluate analytically the integral over k_{y_i} in equation (7.7). Going back to equation (7.1), the integral over k_{y_2} is straightforward to evaluate:

$$\int_{-\infty}^{\infty}\exp\left(-\frac{i\Delta z}{2k}k_{y_2}^2 + ik_{y_2}(y_2'' - y_2')\right)dk_{y_2} = \sqrt{\frac{2\pi k}{i\Delta z}}\exp\left(\frac{ik}{2\Delta z}(y_2'' - y_2')^2\right) \tag{7.9}$$

However, doing such integrals first makes the rhombus overlap integrals difficult to evaluate exactly. For a single interface the expression for the overlap is similar to that given in equation (7.5) but the exponent is now

$$\frac{ik}{2\Delta z}[(y_2'' - y_2')^2 + (y_1' - y_1)^2]$$

The dependence on k_{y_1} and k_{y_2} having been "Fourier transformed" into dependence on y_1 and y_2''. If limiting cases are considered, then some progress is possible. Specifically, if $a \sim \Delta\theta^p$ as $\Delta\theta \to 0$ with $p > 2$, then using γ and Δ as the variables of integration (as defined in appendix E) and taking the limit $\Delta\theta \to 0$ allows the exponential to be brought outside the integrals, which then separate to give a product. Each integrand is a cosine.

The result for 2 stages, 1 interface is:

$$\bar{J}(n_2,y_2'',j_2;n_1,y_1,j_1)$$
$$= FE\left(\frac{a^2}{\theta}\right)\left(\frac{2\sin\frac{1}{2}n_2\pi}{n_2\pi}\right)\left(\frac{2\sin\frac{1}{2}n_1\pi}{n_1\pi}\right)\exp\left(\frac{ik}{2\Delta z}[(y_2'' - y_{c2}')^2 + (y_{c2}' - y_1)^2]\right) \tag{7.10}$$

where

$$F = \left(\sqrt{\frac{2\pi k}{i\Delta z}}\right)^2 \left(\sqrt{\frac{2}{a}}\right)^2 \left(\frac{1}{\sqrt{2\pi}}\right)^4$$

$$E = \exp\left[ik(2\Delta z) - \frac{i\Delta z}{2k}\left(\frac{\pi}{a}\right)^2 (n_1^2 + n_2^2)\right]$$

It is believed that this limit is likely to correspond to the vakonomic case.

Conversely, if $\frac{\Delta\theta}{a} \to 0$ as $\Delta\theta \to 0$, then using γ and Δ as variables of integration and taking the limit $\Delta\theta \to 0$ allows the integral over γ to be evaluated to a (lowest order) stationary phase approximation. The integrand of the remaining integral (over Δ) is then a product of two cosines.

The result for 2 stages, 1 interface is:

$$\bar{J}(n_2, y_2'', j_2; n_1, y_1, j_1) = F\,E\left(\frac{a^2}{\theta}\right)\frac{1}{2}\delta_{n_1 n_2}\exp\left(\frac{ik}{2(2\Delta z)}(y_2'' - y_1)^2\right) \tag{7.11}$$

where F and E are the same as in equation (7.10) and $\delta_{n_1 n_2}$ is a Kronecker delta function. It is believed that this limit is likely to correspond to the ordinary nonholonomic case.

Extending the calculations to 3 stages (2 interfaces) suggests that the N-stage (N-1 interface) results will take the form

"Vakonomic"

$$\begin{aligned}
&\bar{J}(n_N, y_N^{(N)}, j_N; n_1, y_1, j_1) \\
&= \sum_{j_2=-\infty}^{\infty} \cdots \sum_{j_{N-1}=-\infty}^{\infty} F_N \left(\frac{a^2}{\theta}\right)^{(N-1)} \left(\frac{2\sin\frac{1}{2}n_N\pi}{n_N\pi}\right)\left(\frac{2\sin\frac{1}{2}n_1\pi}{n_1\pi}\right) \\
&\quad \times \exp\left(\frac{ik}{2\Delta z}\left[\left(y_N^{(N)} - y_{c(N-1)}^{(N-1)}\right)^2 + \left(y_{c1}' - y_1\right)^2\right]\right) \\
&\quad \times \left[\prod_{l=2}^{N-1} \exp\left(\frac{ik}{2\Delta z}\left(y_c^{(l)}{}_l - y_{c(l-1)}^{(l-1)}\right)^2\right)\right] \\
&\quad \times [\exp(ik(N\Delta z))]\left[\exp\left(-\frac{i\Delta z}{2k}\left(\frac{\pi}{a}\right)^2 (n_1^2 + n_N^2)\right)\right] \\
&\quad \times \left[\sum_{m=0}^{\infty} \frac{1}{(2m-1)^2} \exp\left(-\frac{i\Delta z}{2k}\left(\frac{\pi}{a}\right)^2 (2m-1)^2\right)\right]^{(N-2)}
\end{aligned} \tag{7.12}$$

"Nonholonomic"

$$\begin{aligned}
&\bar{J}(n_N, y_N^{(N)}, j_N; n_1, y_1, j_1) \\
&= \sum_{j_2=-\infty}^{\infty} \cdots \sum_{j_{N-1}=-\infty}^{\infty} F_N \left(\frac{a^2}{\theta}\right)^{(N-1)} \left(\frac{1}{2}\right)^{(N-1)}
\end{aligned}$$

$$\times \exp\left(\frac{ik}{2(N\Delta z)}\left[\left(y_N^{(N)} - y_{cN}^{(N-1)}\right) + \sum_{l=2}^{N-1}\left({y_c^{(l)}}_l - {y_c^{(l-1)}}_l\right) + (y'_{c1} - y_1)\right]\right)$$

$$\times[\exp(ik(N\Delta z))]\left[\exp\left(-\frac{i\Delta z}{2k}\left(\frac{\pi}{a}\right)^2 (n_1^2 + n_N^2)\right)\right]$$

$$\times\left[\exp\left(-\frac{i(N-2)\Delta z}{2k}\left(\frac{\pi}{a}\right)^2 n_1^2\right)\right]\delta_{n_1 n_N} \tag{7.13}$$

in both expressions

$$F_N = \left(\sqrt{\frac{2\pi k}{i\Delta z}}\right)^N \left(\sqrt{\frac{2}{a}}\right)^{2(N-1)} \left(\frac{1}{\sqrt{2\pi}}\right)^{2N} \tag{7.14}$$

and

$${y_c^{(l)}}_l = {x_{c(l+1)}^{(l)}} \csc\Delta\theta - {x_c^{(l)}}_l \cot\Delta\theta \tag{7.15}$$

$${y_c^{(l-1)}}_l = {x_c^{(l-1)}}_l \cot\Delta\theta - {x_c^{(l-1)}}_l \csc\Delta\theta \tag{7.16}$$

$${x_c^{(l)}}_l = j_l a \tag{7.17}$$

$${x_c^{(l-1)}}_l = j_l a \tag{7.18}$$

The expression (7.12) has the sum of the squares of the infinitesimal displacements in the exponent. This is like the path length, which is what would be expected in a vakonomic expression. The "opposite" limiting case (7.13) ($\frac{\Delta\theta}{a} \to 0$) is expected to be nonholonomic . It has the square of the sum of the infinitesimal displacements in its exponent. There are intermediate cases in addition to these extremes.

Although the expressions have been written for the general n-stage case, it is not certain that they will give the correct limit for a finite propagation. Terms have been neglected which are small for a single stage but which might combine to become significant in such a limit. It is therefore desirable to perform some numerical calculations using a general formula, with the hope of checking these special cases in the appropriate limits.

For the calculations, the continuous variable is discretised and a "matrix" based upon formula (7.5) used to perform multiple stage evolution of an initial wavefunction (Typically a gaussian in k_y and specified n and j). Quantities such as the overlap with a final wavefunction or the distribution of probability between lanes (j) can then be calculated. An important check on the computations is that probability is conserved when all "channels" are considered. The computer resources required to perform the calculations increase rapidly with the ranges of the indices of the matrix, so (fairly drastic) truncation of the sums is inevitable. This means that probability is "lost" into the neglected "channels".

To model the continuously rotating case the value of $\Delta\theta$ should be chosen to be small. However, taking $\Delta\theta = \frac{\pi}{2}$ offers the possibility that only a "unit cell" of a small number of squares (i.e. rhombuses with $\theta = \frac{\pi}{2}$) need be considered if the initial wavefunction is chosen suitably. After two stages the constraint planes return to their initial orientation, so this is the natural interval to consider in the z-direction. Reducing the sum over the "lane index" (j) allows the range of the others to be increased, for the same "computational load". Also, the symmetry of the $\Delta\theta = \frac{\pi}{2}$ case means that it is possible to construct initial wavefunctions which should be invariant under propagation. Some of these contain only low modes. These simple eigenvectors of the finite matrix have eigenvalues of modulus unity even when the "truncated" matrix is considered. Numerical calculation of the eigenvectors confirms the presence of these and also shows further, more complicated, eigenvectors with eigenvalues of modulus approximately unity.

This ($\Delta\theta = \frac{\pi}{2}$) example shows that, even in an artificially simple situation, the restrictions on the number of modes required to make the calculations practical are a severe limitation. Consequently a meaningful numerical comparison with the limiting case results is excessively ambitious. A new method which is calculationally more straightforward is required. An approach to this problem is presented in the next chapter.

7.4 Summary

The idea of the approach in this chapter is to start with a mode of the "wave guide", propagate it through the model constrained system, and then calculate the (transition) amplitude to end up "scattered" into a given final mode. The last step is achieved by taking the "overlap integral" with the specified final mode. In fact, one can begin with a general (periodic period $2a$) function and carry out the process on each of its Fourier components (the Fourier coefficients can be found by performing "overlap integrals").

The single stage case is considered first. This takes a simple form, once evaluated, because there is no opportunity for "scattering" into other modes, so the result is zero unless the specified final mode is the same as the initial mode. It is at the interface between two stages that "scattering" takes place (provided there is a non-zero angle of rotation between them).

Building on the result for the single stage, the two stage case is given, to indicate the principle behind "composition" of stages. The expressions are now more complicated. The "overlap integral" on the interface plane between two stages rotated relative to each other,

is now required. This may be written as a sum of integrals over the individual rhombus cells. The edges of the cells are defined by the constraint strips in the two stages meeting at the interface.

The only purpose of the stacks of strips is to enforce the constraint, so it is the average over all suitable stacks that should really be used. Introducing "shifts" allows such an "average over shifts" to be performed when convenient.

In order to make the composition of stages more systematic, it is desirable to define a quantity "J" (equation (7.6)). The simplest form (of "J") is just the smallest repeating unit in a many stage propagation. For composition of stages, it is desirable to evaluate analytically as many of the sums and integrals as possible. If the final formula is to be used for numerical work, then evaluating the integrals (i.e. the "rhombus integral" and the integral over k_{y_i}) analytically would be particularly beneficial, as "discretization" errors would be avoided. It turns out that the option of performing both types of integrals exactly is not available. The method pursued is to do the integral over k_{y_i} exactly and then approximate the "rhombus integral" in the conjectured vakonomic and nonholonomic limiting cases. Since this is the "systematic quantity", "J" (equation (7.6)), it is straightforward to make a formal generalisation to an arbitrary number of stages. To go further and demonstrate that the errors do indeed remain small under "composition" is quite an involved process. Even if this were completed, some sort of check on the result would still be required. It seems sensible to verify that a suitable method of checking the result is available first. A numerical (computational) investigation using a fundamental form of the equation seems to be a possibility, until one attempts to obtain results of quality sufficient for verification purposes.

If a different approach gave results consistent with those obtained in this chapter, then confidence in the results would be increased. A different approach is introduced in the next chapter.

Chapter 8

Phase screens

8.1 Introduction

This chapter introduces a new model and provides an investigation of its behaviour. In the general case the constraint is only applied approximately. However, scope for violation of the constraints is restricted when the limit conjectured to correspond to the nonholonomic case is taken. It is worth considering this model because it is the nonholonomic case which is of particular interest. It is hoped that an advantage of this model will be that the "composition" of stages and formulation of a path integral will be more straightforward than for the "mirror planes" model introduced and studied in chapter 6 (and also investigated in chapter 7).

8.2 Preliminaries

A "phase screen", composed of (parallel) strips of phase plate which introduce a phase shift of π (i.e. a sign change) alternating with strips which give no phase shift (figure 8.1), modifies an incident wavefunction so that directly after passing through the screen it is zero along the boundaries between the two types of strip. If the strips are of equal width (and infinite length) and the incident wavefunction is uniform (constant across the screen) then this pattern of zeros persists indefinitely. This is not true for a general incident wavefunction but will continue to hold to a good approximation for $\Delta z \ll ka^2$. The paraxial approximation is used: Δz is the distance along the "axial direction" ("axis of propagation"). The width of the strips is a and k is the magnitude of the wave-vector for the incident plane-wave. If two identical phase screens of this type are placed next to each other with their strips aligned, then the net effect is zero. However, if they are separated

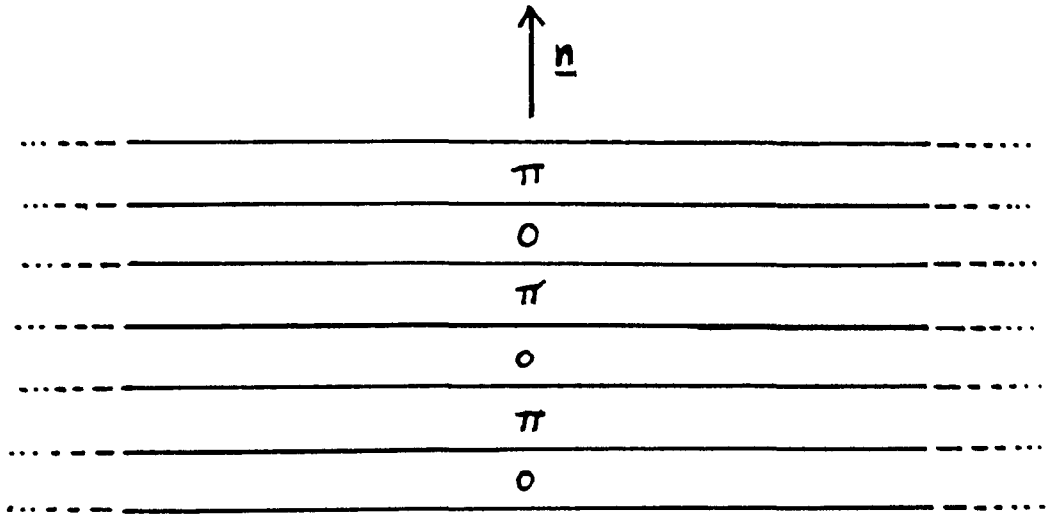


Figure 6.1: Phase plate strips in x-y plane

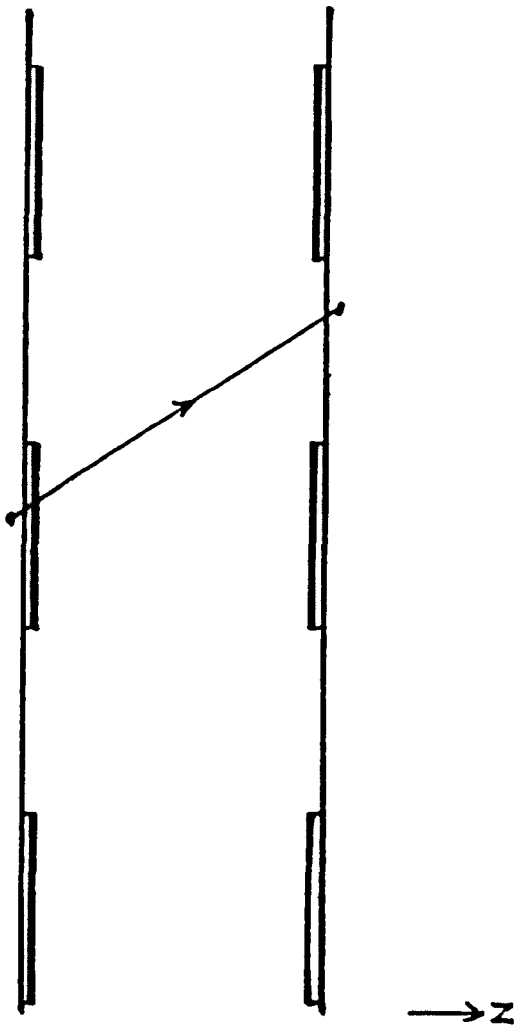


Figure 6.2: Section through a single stage, showing phase screen pair

by a small (compared to ka^2) distance in the direction of propagation (figure 8.2), then, for $a \to 0$, the resulting system (appendix G) may be compared to a single stage of the "mirror planes" model (chapter 5) in the "nonholonomic" ($\frac{\Delta z}{a} \to 0$) limit. The mirror planes ensure that the wavefunction is zero throughout the stage whereas the phase screens enforce the constraint at the ends of the stage only. Between the phase screens is "free space", but the restriction on the length of the stage means that the wavefunction has no chance to violate the constraints significantly.

The comparison between "modes" (i.e. an implementation of the "mirror-planes" model relevant for this comparison) and "phase screens" is made more quantitative in appendix H. An incident wavefunction sinusoidally varying in the transverse direction (i.e. transverse to the direction of propagation) will be "scattered" by passage through a stage into similar wavefunctions of a range of frequencies. The long wavelengths ($\frac{\lambda}{a} \gg 1$) are of most interest, as it is hoped that the general large scale features will be the same. The short wavelengths are "noise" in the sense that they are likely to reflect the particular way in which the constraint is applied. So, in appendix H, the overlap integral for initial and final "long sinusoids" is calculated for both modes and phase screens. The result is that agreement is best for long wave-lengths, just as one would hope.

8.3 A simple case

For phase screens it is natural to consider what happens to a uniform plane-wave when it is incident upon a single stage. After propagation through the stage, the overlap integral with another constant amplitude wavefunction is calculated. For "modes" this is not such an obvious thing to consider, but in this simple case both results are the same, i.e.

$$\frac{1}{2}\left(\frac{4}{\pi}\right)^2 e^{ik\Delta z} \sum_{n=1}^{\infty}{}^{odd} \frac{1}{n^2} e^{-\frac{i\Delta z}{2k}\left(\frac{n\pi}{a}\right)^2}$$

The function

$$f(Z) = \sum_{n=1}^{\infty}{}^{odd} \frac{1}{n^2} e^{iZn^2} \tag{8.1}$$

appearing in this expression is a fractal function since its derivative,

$$\frac{df}{dZ} = i \sum_{n=1}^{\infty}{}^{odd} e^{iZn^2} \tag{8.2}$$

is a divergent sum [6] (whilst the original sum converges). The real and imaginary parts of the function (i.e. $\Re(f)$ and $\Im(f)$) are periodic with period 2π (the function is periodic,

period 2π, since each term in the sum is periodic, period 2π). To determine the period of the modulus $|f(Z)|$ it is required to find the smallest θ such that

$$\left|\sum_{n=1}^{\infty}{}^{odd}\frac{1}{n^2}e^{i(x+\theta)n^2}\right| = \left|\sum_{n=1}^{\infty}{}^{odd}\frac{1}{n^2}e^{ixn^2}\right| \tag{8.3}$$

Equivalently the condition is

$$\sum_{n=1}^{\infty}{}^{odd}\frac{1}{n^2}e^{i(x+\theta)n^2} = e^{2\pi i\phi}\sum_{n=1}^{\infty}{}^{odd}\frac{1}{n^2}e^{ixn^2} \tag{8.4}$$

where ϕ may be chosen in the range $0 \leq \phi < 1$.

This will be true if

$$e^{in^2\theta} = e^{2\pi i\phi} \tag{8.5}$$

for all odd (positive) n.

This condition is expected to be necessary as well as sufficient. If t is defined by $t = \frac{\theta}{2\pi}$ then the graphical interpretation of the condition (8.5) is that (on a graph of y against t figure 8.3) there must be a value of t such that plotting $y = \{n^2t\}$ (where {x} denotes the fractional part of x) yields intersections with the line $y = \phi$ (for some ϕ to be chosen in the interval $[0,1)$) for all odd n, i.e. there is a multiple intersection at the point (t,ϕ). Plotting just the $n = 1$ and $n = 3$ cases (figure 8.3) shows that this intersection cannot be at a smaller value of t than the intersection between the lines $y = t$ and $y = 9t - 1$, i.e. $t = \frac{1}{8}$. Evaluating $f(Z_0 + \frac{\pi}{4})$ shows that $|f(Z)|$ is indeed periodic with period $\frac{\pi}{4}$

$$\begin{aligned} f\left(Z_0+\frac{\pi}{4}\right) &= \sum_{m=0}^{\infty}\frac{1}{(2m+1)^2}e^{i(Z_0+\frac{\pi}{4})(2m+1)^2} \\ &= e^{i\frac{\pi}{4}}\sum_{m=0}^{\infty}\left(\frac{1}{(2m+1)^2}e^{iZ_0(2m+1)^2}\right)e^{2\pi i\left(\frac{m(m+1)}{2}\right)} \\ &= e^{i\frac{\pi}{4}}f(Z_0) \end{aligned} \tag{8.6}$$

since $\frac{m(m+1)}{2}$ is an integer (either m or $m+1$ must be even)

so

$$\left|f\left(Z_0+\frac{\pi}{4}\right)\right| = |f(Z_0)| \tag{8.7}$$

as required.

The maximum value of $|f(Z)|$ may be obtained by substituting $Z = 0$, i.e.

$$\begin{aligned} f(0) &= \sum_{n=1}^{\infty}\frac{1}{n^2} - \sum_{j=1}^{\infty}\frac{1}{(2j)^2} \\ &= \left(1-\frac{1}{4}\right)\zeta(2) \end{aligned} \tag{8.8}$$

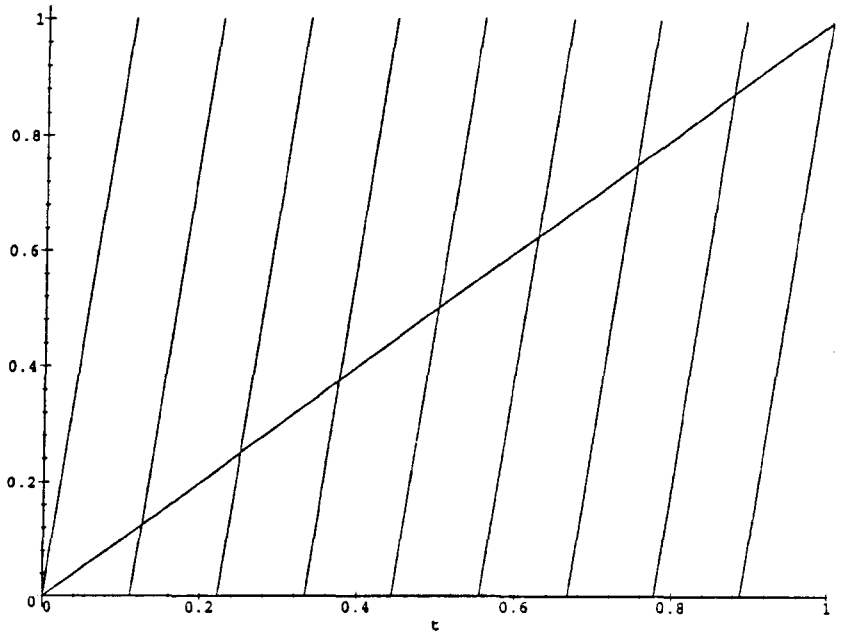


Figure 8.3: Graph of y against t

The result $\zeta(2) = \frac{\pi^2}{6}$ for the Riemann zeta function of 2, gives $f(0) = \frac{\pi^2}{8}$. Using this result to evaluate the overlap integral at $Z = 0$ gives 1 as expected.

Evaluating $f(Z)$ at $Z = \frac{\pi}{8}$, i.e. half the period of the modulus of the function, gives

$$f\left(\frac{\pi}{8}\right) = e^{i\frac{\pi}{8}} \sum_{k=-\infty}^{\infty} \frac{(-1)^k}{(4k+1)^2} \tag{8.9}$$

so $|f(\frac{\pi}{8})| \approx 0.87$

The graph 8.4 (the solid line is $\Re(f)$) shows that this does not in fact give the minimum value for $|f(Z)|$, but that the true minimum value is close to this. The conclusion is that, for a plane-wave entering a single stage, most of the energy remains in the constant component. Only about 30% of the total is ever in the other components.

8.4 Composition of stages

8.4.1 Introduction

The propagator for phase screens is (for $z_2 > z_1$)

$$K(\underline{r}_2, z_2; \underline{r}_1, z_1) = \int_{\underline{r}_1, z_1}^{\underline{r}_2, z_2} e^{i\int_{z_1}^{z_2}\left(\frac{\mu}{2}\dot{\underline{r}}^2 + \Phi(\underline{r}, z)\right)dz} d^{\infty}\underline{r}(z) \tag{8.10}$$

$$\mu = k \qquad \text{(optics)} \tag{8.11}$$

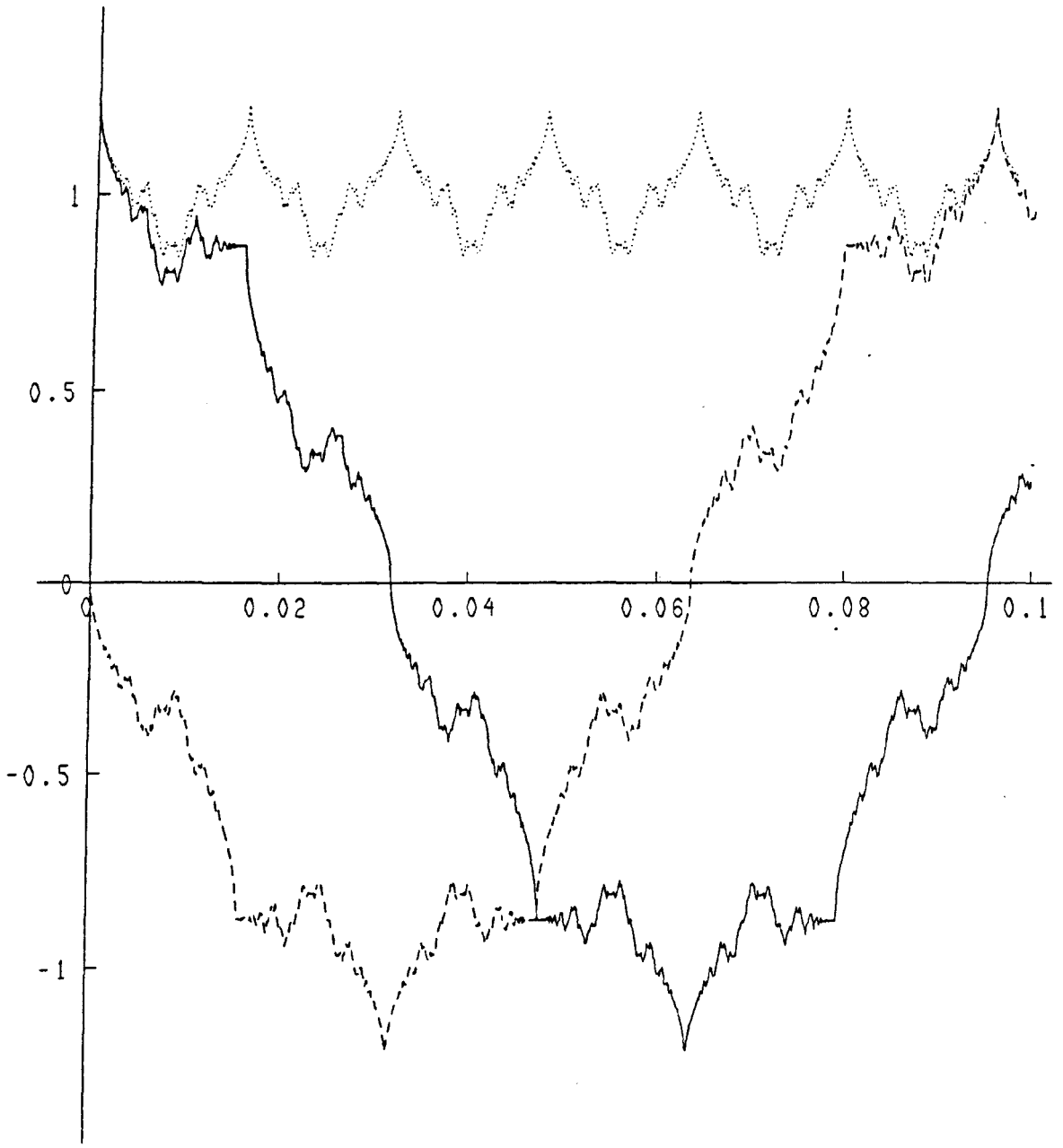


Figure 6.4: $\Re(f), \Im(f), |f|$ against Z

or

$$\mu = \frac{mc}{\hbar} \qquad \text{(mechanics)} \tag{8.12}$$

The "phase function" $\Phi(\underline{r}, z)$ acts only at the interfaces (labelled by n) between stages (of length Δz) i.e.

$$\Phi(\underline{r}, z) = \sum_n \phi_n(\underline{r})\delta(z - n\Delta z) \tag{8.13}$$

where the function $\phi(\underline{r})$ is obtained by adding the phases of two adjacent phase screens with the second rotated slightly relative to the first. This gives a grid of rhombuses with values "modulo 2π" of either 0 or π. Rhombuses with an edge in common have different values.

Each stage (separated parallel phase screen pair) is invariant under translation parallel to the strips forming the phase screens but not under "shifts" in the orthogonal direction. If the shift associated with the i^{th} stage is α_i then the propagator depends upon all such shifts, i.e. α_i for all i. This is undesirable since the values of the shifts are arbitrary. It is natural to average over all the shifts but the averaged propagator $\langle K \rangle_\alpha$ may reduce to something "trivial". The quantity $K^*(\underline{r}'_2, z_2; \underline{r}'_1, z_1)K(\underline{r}''_2, z_2; \underline{r}''_1, z_1)$ (abbreviated to $K^*(2';1')K(2'';1'')$) should be less dependent upon the values of the shifts. Its average $\langle K^*(2';1')K(2'';1'')\rangle_\alpha$ will be more physically meaningful. The intensity (from a "point source") is obtained when the initial points coincide ($1'' = 1'$) and the final points are the same as well ($2'' = 2'$).

Evaluation of the path integrals will be simplified if averaging is carried out first.

8.4.2 Averaging over shifts

The propagator is proportional to

$$\lim_{N\to\infty} \int d^2\underline{r}_1 \ldots \int d^2\underline{r}_{N-1} \prod_{j=1}^{N} e^{i\nu(\underline{r}_j - \underline{r}_{j-1})^2} e^{i\pi\left[\frac{x_{j-1}-\alpha_j}{a}\right]} e^{i\pi\left[\frac{x_j-\alpha_j}{a}\right]}$$

where $\nu = \frac{\mu}{2}\left(\frac{z_2-z_1}{N}\right)^{-1}$, $x_j = \underline{r}_j.\underline{n}_j$ (i.e. each stage has a "local" set of coordinates, x_j being the notation used to represent the coordinate in the $\underline{n}$ direction for "stage j") and [] means "integer part of" (largest integer not greater than)

So averaging over shifts involves evaluating expressions of the form

$$\langle e^{i\pi\left[\frac{x_A-\alpha}{a}\right]} e^{i\pi\left[\frac{x_B-\alpha}{a}\right]} \rangle = \frac{1}{2a}\int_{\alpha=0}^{2a} f_{sign}(x_A - \alpha; a) f_{sign}(x_B - \alpha; a) d\alpha \tag{8.14}$$

since the quantity to be averaged is periodic and $f_{sign}(x;a) = \exp\left(i\pi\left[\frac{x}{a}\right]\right)$

If $f_{sign}(x;a)$ is represented by its Fourier series as in appendix G. Then performing the

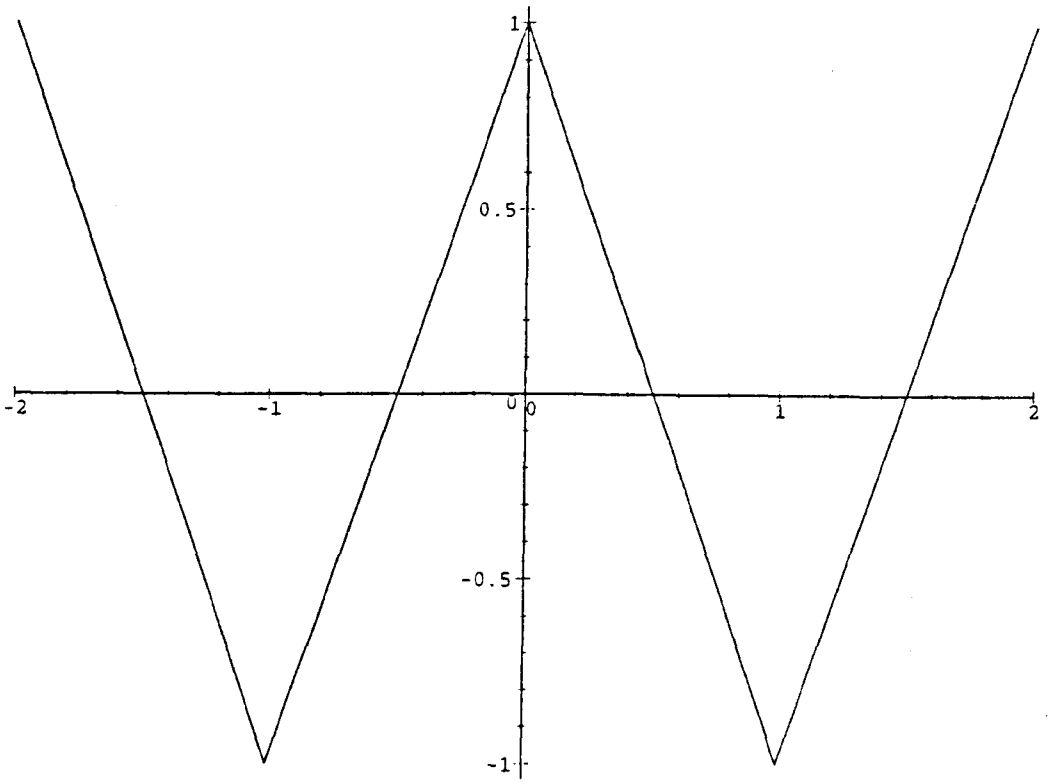


Figure 8.5: Graph of $\langle e^{i(\phi_A+\phi_B)}\rangle$ against ΔX

integration gives

$$\langle e^{i(\phi_A+\phi_B)}\rangle = \frac{8}{\pi^2}\sum_{n\overset{odd}{=}1}^{\infty}\frac{1}{n^2}\cos\left(\frac{n\pi}{a}(x_B-x_A)\right) \tag{8.15}$$

Alternatively, proceeding directly from the graphs of the functions gives

$$\langle e^{i(\phi_A+\phi_B)}\rangle = 1-4\left|\frac{1}{2}(\Delta X-1)-\left[\frac{1}{2}(\Delta X-1)\right]-\frac{1}{2}\right| \tag{8.16}$$

where $\Delta X = \frac{x_B-x_A}{a}$

It is straightforward to verify that the Fourier series representation of this function is indeed given by equation (8.15). Averaging over shifts has removed dependence on the average position $\frac{1}{2}(x_A+x_B)$, leaving only dependence on the difference x_A-x_B. The graph is shown in figure 8.5.

Taking the limit $\frac{a^2}{\epsilon}\to\infty$ should allow the factors $\langle e^{i(\phi_A+\phi_B)}\rangle$ to be approximated by $1-2|\Delta X|$ inside the path integral (since, under these circumstances, it is anticipated that $\Delta X = \frac{\Delta x}{a}$ is small, i.e. $\Delta X \ll 1$). The explanation for this assertion is that, inside the path integral, the path segments are expected to satisfy $|\Delta x|\sim\epsilon^{\frac{1}{2}}$ as $\epsilon\to 0$ [10]. Consequently $\frac{a^2}{\epsilon}\to\infty$ means that $\frac{\Delta x}{a}\to 0$ as $\epsilon\to 0$ since $\frac{\Delta x}{a} = \frac{\Delta x}{\sqrt{\epsilon}}/\sqrt{\frac{a^2}{\epsilon}}$. In the limit $\frac{a^2}{\epsilon}\to\infty$ (which is expected to ensure that the system is in the nonholonomic regime). The constraint is applied to a good approximation, as required (section 8.2).

Making the further approxination $1 - \frac{2}{a}|\Delta x| \approx e^{-\frac{2}{a}|\Delta x|}$ inside the path integral (using the same justification), leads to a simple expression for the averaged propagator in the limit $\frac{a^2}{\epsilon} \to \infty$, i.e.

$$\langle K(\underline{r}_2, z_2; \underline{r}_1, z_1)\rangle = \int_{\underline{r}_1, z_1}^{\underline{r}_2, z_2} e^{\int_{z_1}^{z_2}\left(\frac{i\mu}{2}\dot{\underline{r}}^2 - \frac{2}{a}|\dot{\underline{r}}.\underline{n}(z)|\right)dz} d^\infty \underline{r}(z) \tag{8.17}$$

where Δx has been replaced by $\dot{x}dz$ in the "continuum limit"
($\dot{x} = \dot{\underline{r}}.\underline{n}$ in this section, which makes it similar to a "quasi-coordinate velocity" – section A.3)
For $\langle K^*K\rangle$ the expressions are similar except that there are now four "phase factors" (two from each path) instead of two. So the single stage "sign average" is

$$\langle e^{i(\phi''_A + \phi''_B)} e^{-i(\phi'_A + \phi'_B)}\rangle = \langle e^{i\pi\left[\frac{x''_A - \alpha}{a}\right]} e^{i\pi\left[\frac{x''_B - \alpha}{a}\right]} e^{-i\pi\left[\frac{x'_A - \alpha}{a}\right]} e^{-i\pi\left[\frac{x'_B - \alpha}{a}\right]}\rangle \tag{8.18}$$

It is possible to write these as Fourier series again but (after averaging removes one) there are still three independent variables, so the answer will be complicated, e.g. if $\bar{x}' = \frac{1}{2}(x'_A + x'_B)$, $\bar{x}'' = \frac{1}{2}(x''_A + x''_B)$, $\Delta x' = x'_B - x'_A$, $\Delta x'' = x''_B - x''_A$ then averaging removes dependence on $\frac{1}{2}(\bar{x}' + \bar{x}'')$ giving a function of $\Delta\bar{x} = \bar{x}'' - \bar{x}'$, $\Delta x'$ and $\Delta x''$. It is therefore preferable to consider the limiting case from the beginning. In fact, once this specialisation has been made, the "averaged sign" for a single stage can be written (inside the path integral) in a form similar to that used for $\langle K\rangle$. It is necessary to replace $|\Delta x|$ by the "total non-overlapping length", T, which is a function of $\Delta\bar{x}$, $\Delta x'$ and $\Delta x''$. Projecting the path segments of a single stage onto the direction defined by the normal vector $\underline{n}$ gives two intervals. The paths are said to "overlap" if these intervals coincide over part or all of their length. There are three cases to consider: zero overlap; partial overlap; complete overlap. This is summarized in table 8.1. A graph of $T/|\Delta x'|$ against $\Delta\bar{x}/|\Delta x'|$ for fixed $\Delta x'$ and $\Delta x'' = \frac{1}{2}|\Delta x'|$ is shown in figure 8.6. Neglecting the nonzero-overlap cases as a first approach to evaluating $\langle K^*K\rangle$ means taking the single stage "sign average" to be

$$e^{-\frac{2}{a}(|\Delta x'| + |\Delta x''|)}$$

This leads to the double path integral

$$J(\underline{r}'_2, \underline{r}''_2, z_2; \underline{r}'_1, \underline{r}''_1, z_1) = \int_{\underline{r}''_1, z_1}^{\underline{r}''_2, z_2} \int_{\underline{r}'_1, z_1}^{\underline{r}'_2, z_2} e^{\int_{z_1}^{z_2}\left(\frac{i\mu}{2}(\dot{\underline{r}}''^2 - \dot{\underline{r}}'^2) - \frac{2}{a}(|\dot{\underline{r}}''.\underline{n}| + |\dot{\underline{r}}'.\underline{n}|)\right)dz} d^\infty \underline{r}'(z)\, d^\infty \underline{r}''(z) \tag{8.19}$$

which factorizes

$$J(\underline{r}'_2, \underline{r}''_2, z_2; \underline{r}'_1, \underline{r}''_1, z_1) = \langle K(\underline{r}''_2, z_2; \underline{r}''_1, z_1)\rangle \langle K(\underline{r}'_2, z_2; \underline{r}'_1, z_1)\rangle^* \tag{8.20}$$

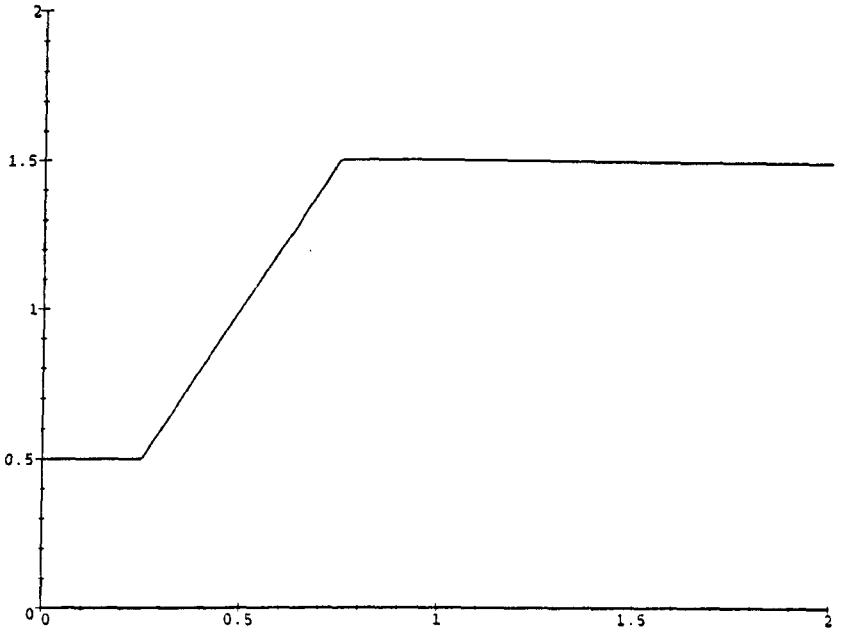


Figure 8.6: Graph of $T/|\Delta x'|$ against $\Delta\bar{x}/|\Delta x'|$ for $\Delta x'' = \frac{1}{2}|\Delta x'|$

Overlap	*Non-overlapping length*
zero	$\|\Delta x'\| + \|\Delta x''\|$
partial	$2\|\Delta\bar{x}\|$
complete	$\|\|\Delta x'\| - \|\Delta x''\|\|$

Table 8.1: T as a function of "overlap"

This is as expected, since the paths were assumed never to interact (i.e. to be independent). In view of this relation, it is interesting to investigate the averaged propagator. First, it is worth checking that probability is conserved under propagation by $\langle K^*K\rangle$.

8.4.3 Conservation of probability

For a single stage (of length $z_B - z_A > 0$)

$$\begin{aligned} & K(\underline{r}''_B, z_B; \underline{r}''_A, z_A) K^*(\underline{r}'_B, z_B; \underline{r}'_A, z_A) \\ &= \sqrt{\frac{\nu}{\pi i}} \exp\left(i\nu(\underline{r}''_B - \underline{r}''_A)^2 + i\pi\left[\frac{x''_A - \alpha}{a}\right] + i\pi\left[\frac{x''_B - \alpha}{a}\right]\right) \\ &\quad \times \sqrt{\frac{\nu}{\pi(-i)}} \exp\left(-i\nu(\underline{r}'_B - \underline{r}'_A)^2 - i\pi\left[\frac{x'_A - \alpha}{a}\right] - i\pi\left[\frac{x'_B - \alpha}{a}\right]\right) \end{aligned} \tag{8.21}$$

setting $\underline{r}''_B = \underline{r}'_B$ and integrating over this final position gives

$$\begin{aligned}
&\int_{-\infty}^{\infty} K(\underline{r}_B, z_B; \underline{r}''_A, z_A) K^*(\underline{r}_B, z_B; \underline{r}'_A, z_A) d^2\underline{r}_B \\
&= \left(\frac{\nu}{\pi}\right)^2 \exp\left(i\nu(\underline{r}''^2_A - \underline{r}'^2_A)\right) \\
&\quad \times \exp\left(i\pi\left[\frac{x''_A - \alpha}{a}\right] + i\pi\left[\frac{x'_A - \alpha}{a}\right]\right) \\
&\quad \times \int_{-\infty}^{\infty} \exp\left(2i\nu(\underline{r}'_A - \underline{r}''_A).\underline{r}_B\right) d^2\underline{r}_B \\
&= F(\underline{r}''_A, \underline{r}'_A, \alpha)\delta^2(\underline{r}'_A - \underline{r}''_A)
\end{aligned} \tag{8.22}$$

(where δ^2 is the standard 2D delta function and the function F is defined by this equation.) Using this in the expression for the final probability gives

$$\begin{aligned}
&\int_{-\infty}^{\infty} \psi(\underline{r}_B, z_B)\psi^*(\underline{r}_B, z_B) d^2\underline{r}_B \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K(\underline{r}_B, z_B; \underline{r}''_A, z_A)\psi(\underline{r}''_A, z_A) d^2\underline{r}''_A \\
&\quad \times \int_{-\infty}^{\infty} K^*(\underline{r}_B, z_B; \underline{r}'_A, z_A)\psi^*(\underline{r}'_A, z_A) d^2\underline{r}'_A d^2\underline{r}_B \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(\underline{r}''_A, \underline{r}'_A, \alpha)\delta^2(\underline{r}'_A - \underline{r}''_A)\psi(\underline{r}''_A, z_A)\psi^*(\underline{r}'_A, z_A) d^2\underline{r}''_A d^2\underline{r}'_A \\
&= \int_{-\infty}^{\infty} \psi(\underline{r}_A, z_A)\psi^*(\underline{r}_A, z_A) d^2\underline{r}_A
\end{aligned} \tag{8.23}$$

since $F(\underline{r}_A, \underline{r}_A, \alpha) = 1$

So probability is conserved for a single stage. This process may be iterated to show that probability is conserved for finite propagation.

8.4.4 The averaged propagator

Although the path integral for the averaged propagator (equation (8.17)) seems fairly simple, it is difficult to evaluate. There is, however, an alternative approach which involves obtaining a differential equation. In the standard case [10, 32] the Schrödinger equation is deduced by using the propagator to propagate a wavefunction for a infinitesimal time interval. A similar approach can be applied in this case but, due to averaging, the quantity "ψ" in the calculation should not be interpreted as a wavefunction in the conventional sense. With this proviso (and using time, t, rather than z)

$$\psi(\underline{r}, t + \delta t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \langle K(\underline{r}, t + \delta t; \underline{q}, t))\psi(\underline{q}, t) d^2\underline{q} \tag{8.24}$$

leads to

$$\delta\psi = -\frac{2}{a}\sqrt{\frac{2i}{\pi\mu}}\psi\sqrt{\delta t} + O(\delta t) \quad \text{as } \delta t \to 0 \tag{8.25}$$

In the standard case the leading order term is of order δt, leading to a differential equation, in this case it is of order $\delta t^{\frac{1}{2}}$ (there is still a term of order δt). The term of order $\delta t^{\frac{1}{2}}$ is a consequence of the modulus sign in the expression for $\langle K \rangle$. It suggests that ψ is non-differentiable (a fractal function). The factor has a negative real part suggesting that ψ will decrease with time.

If the presence of the modulus is the most important feature, then the "1D" (i.e. 1D if the z-direction is not counted) version (which is holonomic since at least 2 space dimensions are required for nonholonomy) should display a similar general behaviour i.e. considering the expression for changing from the fixed end point path integral to the free end point version

$$\int_{x_1,z_1}^{x_2,z_2} e^{\int_{z_1}^{z_2}\left(\frac{i\mu}{2}\dot{x}^2-\frac{2}{a}|\dot{x}|\right)dz} d^{\infty}x(z) = \int_{z_1}^{z_2}\int_{-\infty}^{\infty} e^{\int_{z_1}^{z_2}\left(\frac{i\mu}{2}\dot{x}^2-\frac{2}{a}|\dot{x}|+ib\dot{x}\right)dz} e^{-ib\Delta x} db\, d^{\infty}\dot{x}(z) \quad (8.26)$$

(where μ is defined after equation (8.10) but $\dot{x}$ now represents $\frac{dx}{dz}$, the derivative of a position coordinate, rather than the component of the velocity in the non-constant $\underline{n}$-direction)
The path integral on the right hand side of equation (8.26) may be factorized to give a product of N ordinary integrals ($N \to \infty$) of the type

$$I = \sqrt{\frac{\nu}{\pi i}} \int_{y=-\infty}^{\infty} e^{\left(i\nu y^2-\frac{2}{a}|y|+iby\right)} dy \quad (8.27)$$

where $\nu = \frac{\mu}{2\left(\frac{z_2-z_1}{N}\right)}$ and $\frac{z_2-z_1}{N} \to 0$ to obtain the path integral
Evaluating this integral by splitting the range of integration gives

$$I = e^{\frac{i}{4\nu a^2}(2-a^2b^2)}\left[e^{\frac{b}{\nu a}}\left(\frac{1}{2}-\frac{1}{\sqrt{\pi i}}E_+\right)+e^{-\frac{b}{\nu a}}\left(\frac{1}{2}-\frac{1}{\sqrt{\pi i}}E_-\right)\right] \quad (8.28)$$

where

$$E_{\pm} = E\left(\pm\frac{b}{2\sqrt{\nu}}+\frac{i}{a\sqrt{\nu}}\right) \quad (8.29)$$

and

$$E(z) = \int_0^z e^{it^2} dt \quad (8.30)$$

Expanding this in small quantities when $\nu \to \infty$ and taking the N^{th} power for finite propagation, the dominant factor is

$$\left(1-\frac{2}{a}\sqrt{\frac{i}{\pi\nu}}\right)^N$$

which tends to zero as $N \to \infty$ ($\nu \sim N$ as $N \to \infty$). To evaluate the path integral it would be necessary to perform an infinite integration over b. The expansion will not be valid over the whole of the range of integration, but the result will hold qualitatively provided the contribution to the integral from large values of b is a small fraction of the total.

8.5 Numerical investigation of $\langle K \rangle$

A numerical approach provides an opportunity to investigate single and multiple stage propagation using the exact (periodic) expression for the single stage average. The 1D single-stage averaged propagator ("$\langle K \rangle_1$") depends on the difference between the initial and final coordinates (Δx) only. There is no time dependence because $\underline{n}(t)$ does not appear in the formula. So the composition of three stages takes the form

$$\int_{-\infty}^{\infty} f(x_3 - x_2) \left(\int_{-\infty}^{\infty} f(x_2 - x_1) f(x_1 - x_0) dx_1 \right) dx_2$$

making the substitutions $\Delta_i = x_i - x_{i-1}$, $i = 1, 2, 3$, this becomes

$$\int_{-\infty}^{\infty} f(\Delta_3 - \Delta_2) \left(\int_{-\infty}^{\infty} f(\Delta_2 - \Delta_1) f(\Delta_1) d\Delta_1 \right) d\Delta_2 = f * (f * f) \tag{8.31}$$

where $*$ means "convolution" (i.e. $f * g \equiv \int_{-\infty}^{\infty} f(v-u) g(u)\, du$ which is only equivalent to the definition of convolution with limits of integration 0 and v when all functions involved are zero for negative values of their arguments)
The Fourier transform of this is F^3 where F is the Fourier transform of f. The N stage version of this result allows the path integral to be built up by repeated convolution or by taking the N^{th} power of the Fourier transform, F, and then using an inverse Fourier transform on the result (similar to equation (8.27) and equation (8.26)).

For a single stage propagator with the Fourier series representation of the phase average (equation (8.15)) it is possible to compose two stages analytically using both methods, for a large number of stages a numerical approach is required. The Fourier transform method is most suitable for this. For a single stage the Fourier transform is

$$G(k) = \frac{8}{\pi^2} e^{-\frac{i}{4\nu}k^2} \sum_{\substack{m=1 \\ m}}^{\infty} \frac{1}{m^2} e^{-\frac{i}{4\nu}\left(\frac{m\pi}{a}\right)^2} \cos \frac{m\pi k}{2a\nu} \tag{8.32}$$

Of most interest are $|G(k)|$ and $|G(k)|^N$ (for N stages). The modulus of $G(k)$ is even and periodic in k. Its graph shows a series of peaks (figure 8.7). These sharpen into spikes (figures 8.8, 8.9) as higher powers of $|G(k)|$ are taken. However, the spikes need not have a single peak. Whether they do or not depends upon the form of $|G(k)|$, which in turn depends rather sensitively on the value of the parameter (e.g. $A = \frac{1}{4\nu}\left(\frac{\pi}{a}\right)^2$ if a scaled variable $K = \frac{2a}{\pi}k$ is used). Graphs of $|G(k)|$ for several values of the parameter are shown in figures 8.10–8.15. This means that further averaging is desirable. This is introduced in the next chapter.

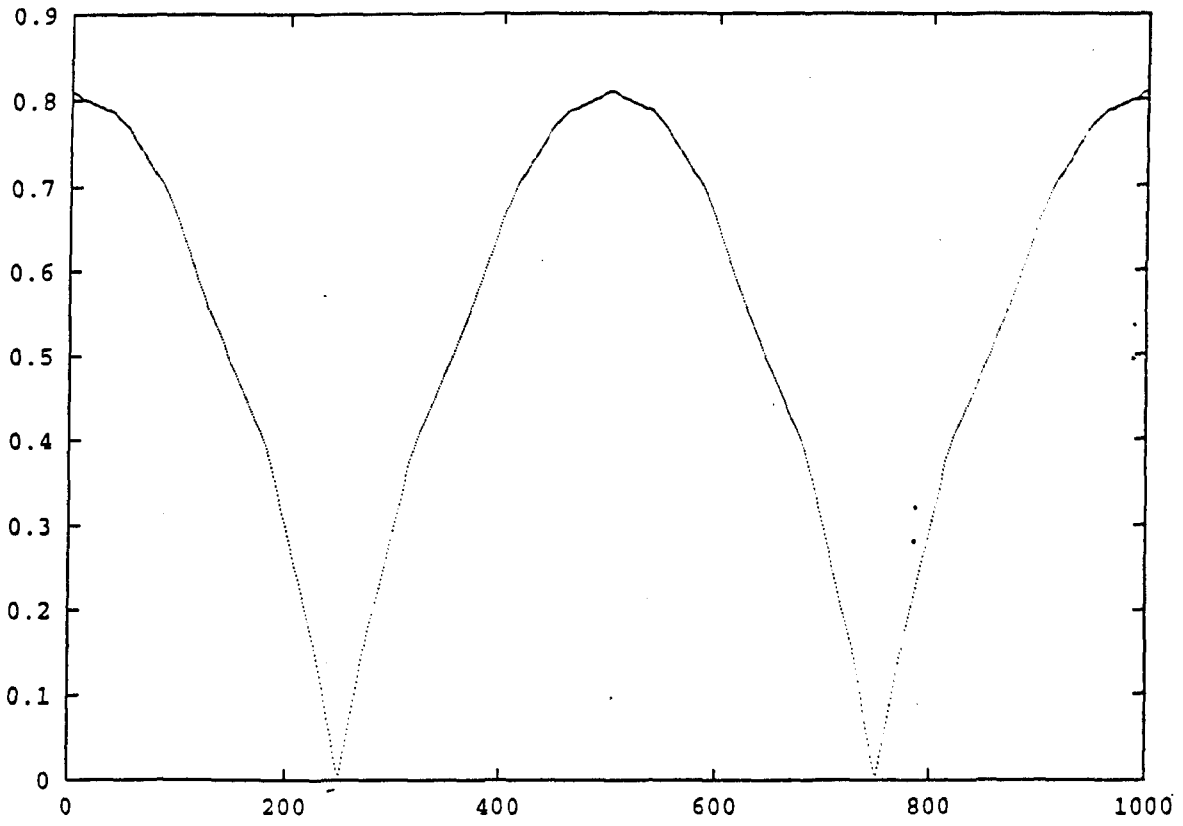


Figure 6.7: Graph of $|G(k)|$ against k with $A = 1$

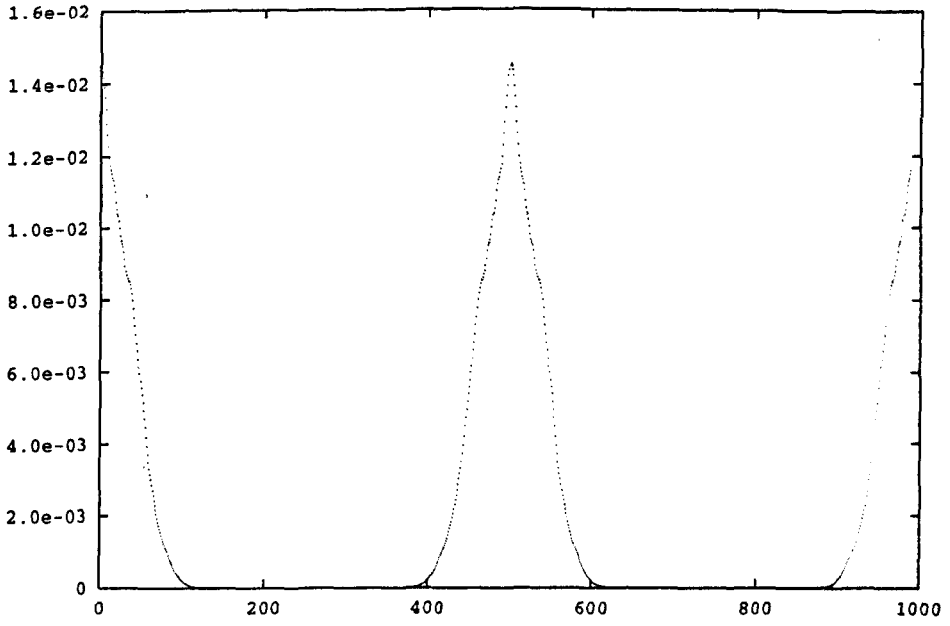


Figure 6.8: Graph of $|G(k)|^N$ against k for $N = 20$ with $A = 1$

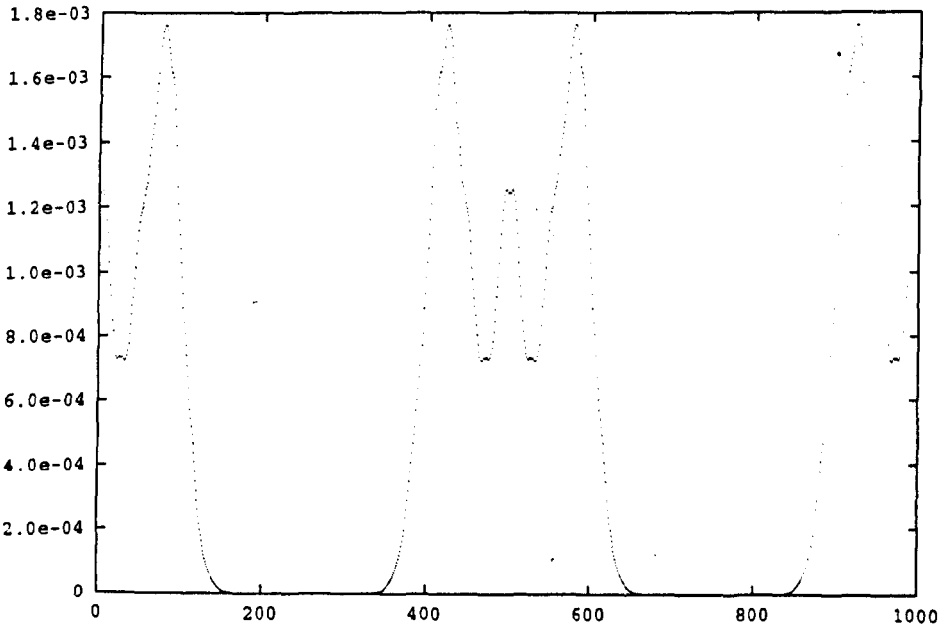


Figure 6.9: Graph of $|G(k)|^N$ against k for $N = 20$ with $A = 1.2$

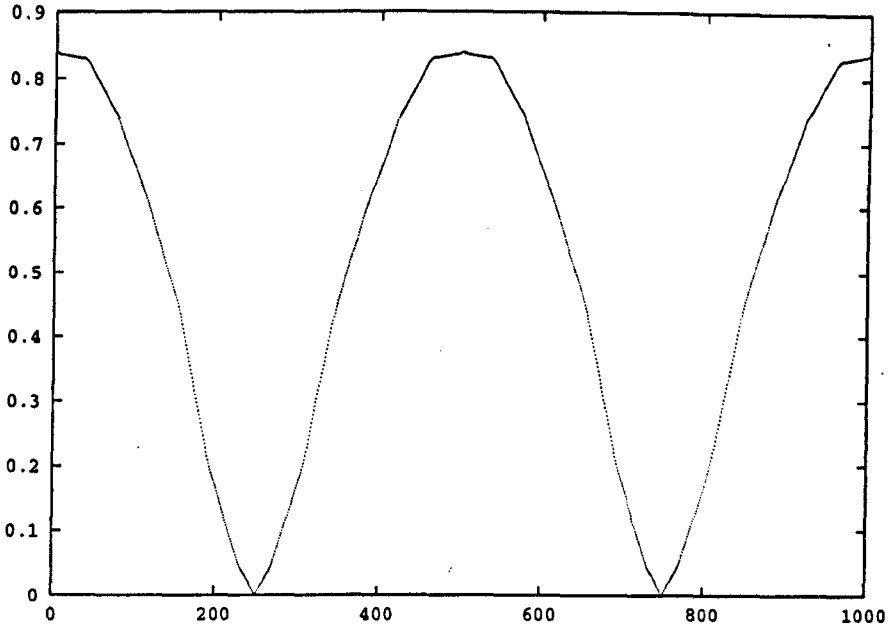


Figure 6.10: Graph of $|G(k)|$ against k with $A = \frac{e}{3}$

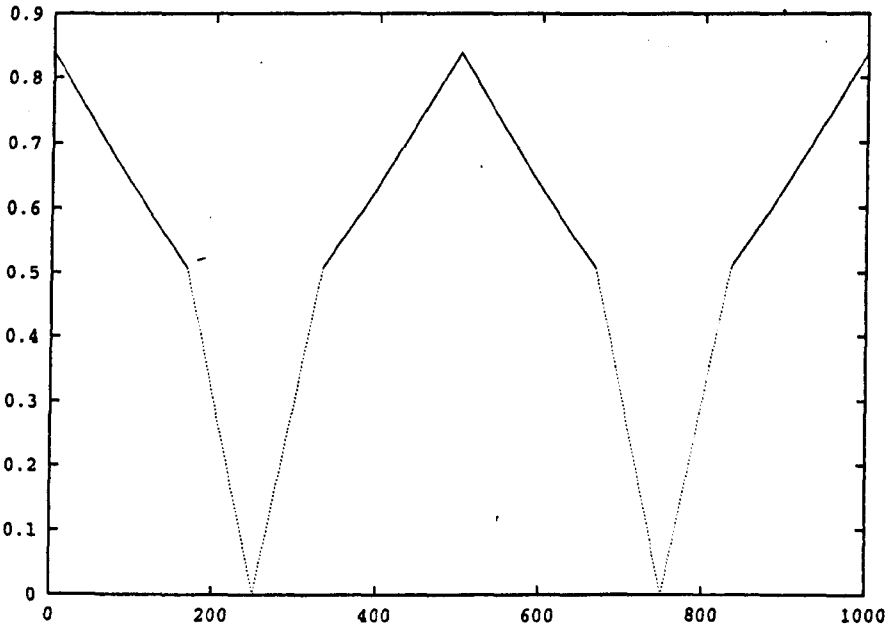


Figure 6.11: Graph of $|G(k)|$ against k with $A = \frac{\pi}{3}$

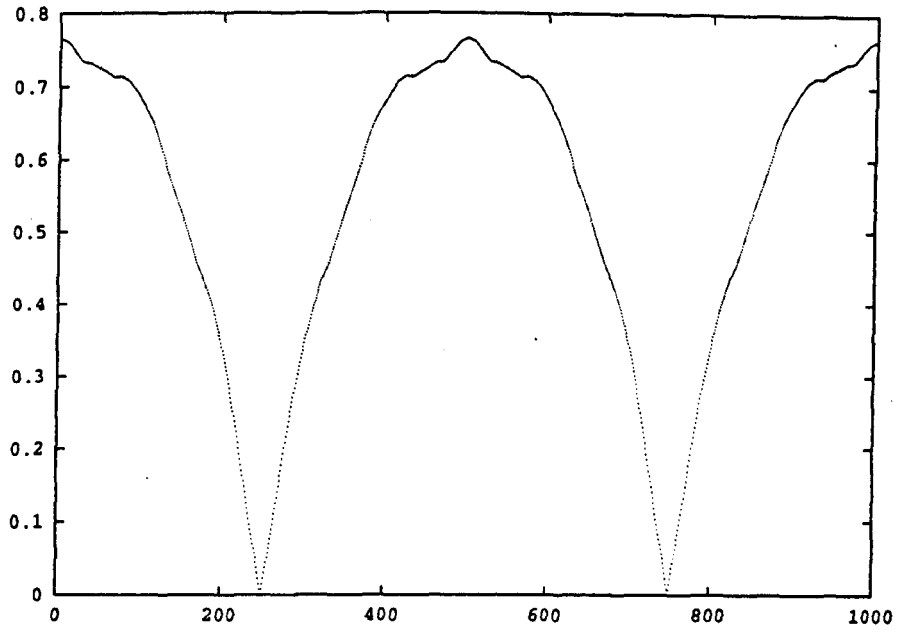


Figure 6.12: Graph of $|G(k)|$ against k with $A = \frac{\pi^2}{9}$

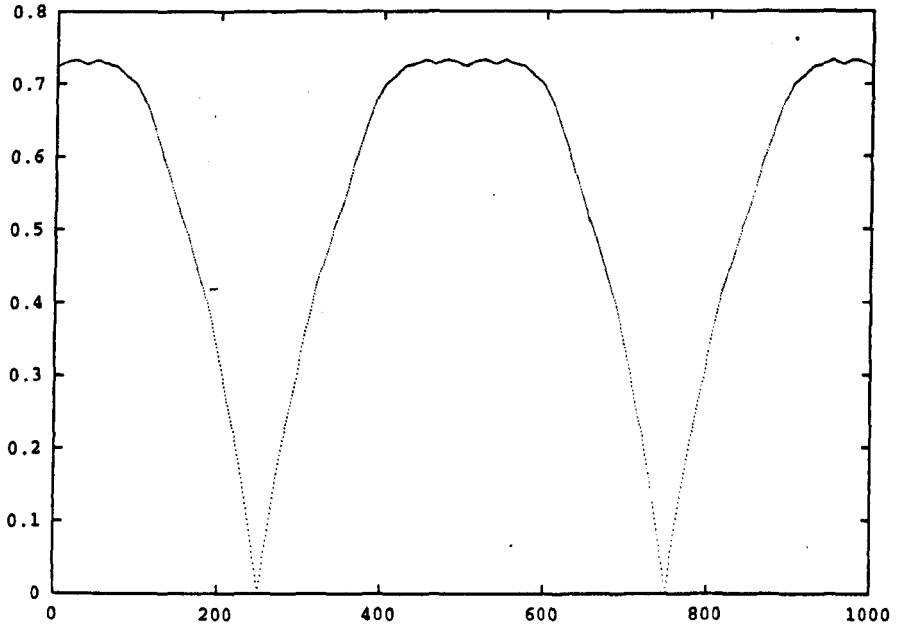


Figure 6.13: Graph of $|G(k)|$ against k with $A = \frac{10}{9}$

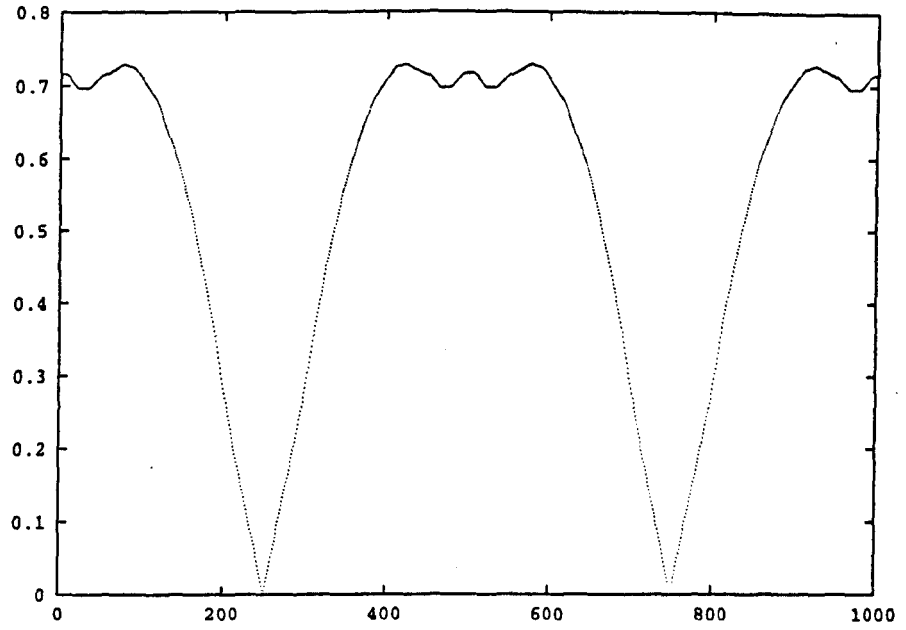


Figure 6.14: Graph of $|G(k)|$ against k with $A = 1.2$

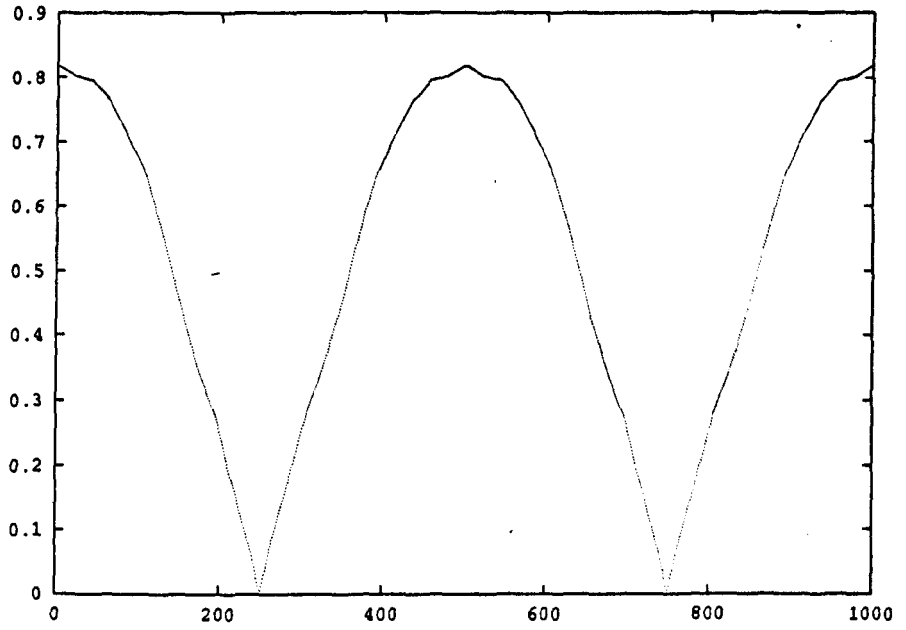


Figure 6.15: Graph of $|G(k)|$ against k with $A = 1.4$

8.6 The classical regime

Phase screens is developed as a way of applying the constraint to the standard path integral. In fact, the standard free space action can be modified to include the "phase function" $\Phi(\underline{r}, z)$ (defined by equation (8.13)). This is just a change in interpretation and it is not really a natural change to make unless the "continuum limit" $\Delta z \rightarrow 0$ has been taken, so that equation (8.13) becomes

$$\Phi_c(\underline{r}, z) = \lim_{\Delta z \rightarrow 0} \sum_n \phi_n(\underline{r}) \delta(z - n\Delta z) \tag{8.33}$$

The fact that $\phi_n(\underline{r})$ depends upon a and $a \rightarrow 0$ as $\Delta z \rightarrow 0$ is part of the reason why this is difficult to evaluate in detail, however, it is fairly clear what the general picture should be. Consequently it should be possible to use this reinterpretation to get some idea of the behaviour of the classical system. So, starting with the modified action

$$S' = \int_{t_1}^{t_2} \left(\frac{m}{2} \dot{\underline{r}}^2 + c\hbar \Phi_c(\underline{r}, ct) \right) dt \tag{8.34}$$

(where mechanics is specifically considered, so $\dot{\underline{r}}$ means $\frac{d\underline{r}}{dt}$ rather than $\frac{d\underline{r}}{dz}$)
and considering $\delta S' = 0$ gives

$$m\ddot{\underline{r}} = -\hbar \frac{\partial \Phi_c}{\partial \underline{r}} \tag{8.35}$$

The discontinuities in the phase function "potential" mean that a particle will be subject to a δ function force if it "hits" the edge of a rhombus.

Moving across the grid of rhombuses, the sign (direction) of the force swaps for alternate boundary lines. A similar pattern occurs when moving over the grid in the other (independent) direction (in 2D space). So a particle will have its motion "out of" a rhombus reversed. However, this potential is only applied at the ends of the stages. There may still (even after taking the limit $\Delta z \rightarrow 0$) be scope for a particle to move more freely than in the mirror planes model (chapter 6). Exactly how the motion is restricted will depend upon the way that $a \rightarrow 0$ as $\Delta z \rightarrow 0$ (section 6.2).

It is only the discontinuous nature of the boundary between the phase-changing and non phase-changing parts of the phase screens which allows their influence to extend to the classical regime. If these discontinuities were "smoothed" then the effect of the phase screens on the classical (i.e. ray) paths would be lost as "$\hbar \rightarrow 0$".

It is not clear, just from these general considerations, whether phase screens meet the requirement of enforcing the constraint correctly in the classical limit and detailed calculations are not straightforward (section 8.4).

8.7 Summary

The phase screens model is introduced. It is noted that the relation between "phase screens" calculations and the "modes" method of calculation for the "mirror planes" model is considered in appendix H. The result is that agreement improves as the "transverse" wave-length (i.e. the wave-length for the $\underline{n}$ direction — shown in figure 8.1) of the wavefunctions becomes long compared to a. In the "mirror planes" model, a is the spacing between adjacent "mirror planes". In the phase screens model the corresponding dimension is the width (i.e. in the $\underline{n}$ direction) of the phase changing strips making up a phase screen.

The agreement between "phase screens" and "modes" is exact for "infinitely long wavelengths" (the constant amplitude wavefunctions case). This simple case is considered explicitly (i.e. rather than as the limit of the more general case in appendix II). The overlap integral is found to be a fractal function of the length of the stage, Δz. This may be a signal of problems with the $\Delta z \to 0$ limit which must be taken in order to pass to the continuum.

It is possible to write down an expression for a path integral. In fact, although the phase screens are constructed separately, it is possible to formally include them in the Lagrangian as a "potential". However, this interpretation should be treated with care as the "potential" will depend upon velocity and will have other non-standard features. This does show that phase screens fit naturally into the structure of path integration.

Although the phase screens can be included in the path integral as it is being "constructed" (section 8.4.2), a suitable interpretation for the "continuum limit", $\epsilon \to 0$, (or alternatively $\Delta z \to 0$) is elusive. In the hope of improving the chance of finding a "continuum form" for the influence the phase screens have in the path integral, "the average over shifts" is carried out: each phase screen pair is shifted through a "periodic distance" (i.e. $2a$) in the $\underline{n}$ direction (i.e. vertically in figure 8.2). This is carried out "inside" (i.e. before) the path integral, so each path segment (also shown in figure 8.2) is considered to be fixed in space.

The result for a single phase screen pair is given but it is still not straightforward to interpret this for the "continuum limit" (i.e. $\epsilon \to 0$). Fortunately, taking the limit believed to correspond to the nonholonomic case allows approximations to be made. These lead to a simple expression for the averaged propagator, $\langle K \rangle$ (and also for the averaged version of K^*K, i.e. $\langle K^*K \rangle$), in the "nonholonomic case". It is desirable to investigate the behaviour of the averaged propagator in order to check its suitability. If its behaviour is suitable then

it will be necessary to check the effect of the approximation in detail, otherwise it will be more profitable to go back and modify the model. From the investigations (sections 8.4.4 and 8.5), the indications are that $\langle K \rangle$ decreases in size as the (time) interval over which propagation takes place becomes finite. The calculations are based on the explicit form obtained for the "nonholonomic limit". In fact, there are general considerations which suggest that the result is not restricted to this case. If $\langle K \rangle$ is evaluated by carrying out the path integral first, then the contribution from each path is a complex number with modulus=1. The path integral is a "sum" of such contributions with different "phases" (arguments). The result depends upon the values of the shifts which are then averaged over. If the average over shifts is carried out first, then the paths with no segments inclined in the $\underline{n}$ direction give a contribution with modulus unity. Most other paths give contributions with modulus smaller than unity — the more steeply the path segments are inclined in the $\underline{n}$ direction (but with allowance for periodicity), and the more segments are so inclined, the smaller the modulus of the contribution. So, if the path integral is now performed, it is a sum of contributions with different phases and different moduli. From this point of view, it is quite plausible that $\langle K \rangle$ might decrease in size in general. At this point it is prudent to consider whether the measure used for the path integration might need to be modified. The fact that probability is conserved (whether averaging is carried out or not — section 8.4.3) suggests that modification is unnecessary.

Chapter 9

Random models

9.1 Introduction

The periodicity of the structures constructed to apply the constraint causes problems in the mirror planes (section 6.3) and the phase screens (section 8.5) models. In this chapter randomness is introduced into both models. In section 9.2 modifications are made to the phase screens model of chapter 8. In section 9.3 analogous changes are introduced into the mirror planes model of chapter 6.

9.2 Phase screens

9.2.1 Introduction

Considering phase screens with equally spaced strips of equal width and then averaging over shifts perpendicular to the strips, causes problems (section 8.5) due to the periodicity of the screens. To avoid such problems, another type of averaging may be introduced. The phase strips and the gaps between them are taken to be of random width (figure 9.1) and averaging is over all possible screens. In fact, the lines marking the edges of the phase strips are of most interest. If these are extended in the direction perpendicular to the phase screens (i.e. in the z direction) to form "planes", then the number of such planes a path segment passes through gives, when multiplied by π, a phase change with the same effect as that due to the phase screen pair. These (segments of) planes will be referred to as "phase counting planes ".

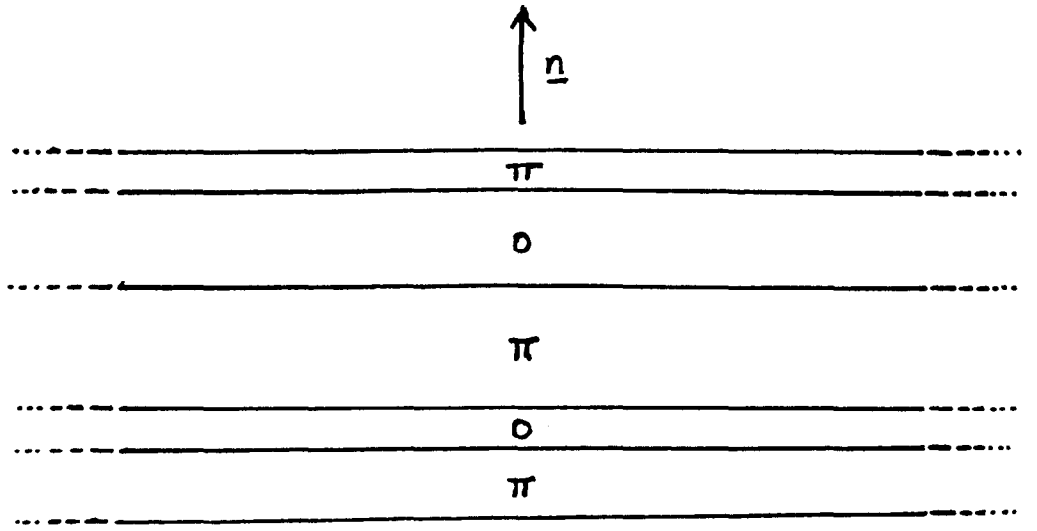


Figure 7.1: Phase plate strips in x-y plane

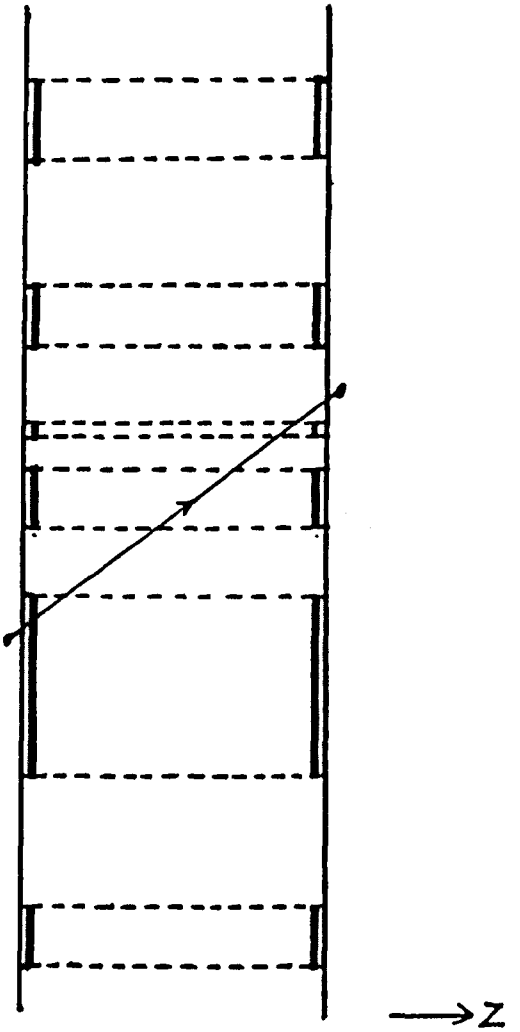


Figure 7.2: Section through a single stage, showing the phase screen pair and the "phase counting planes" (dashed lines)

9.2.2 Preliminaries

Considering a single stage and taking the x-axis in the direction of the "normal vector" (i.e. perpendicular to both the phase strips and the z-axis), the ("phase counting") planes project onto lines in the x-z plane (figure 9.2). To calculate the single stage "sign average" (i.e. the sign change aquired on a path segment between "A" and "B" averaged over the ensemble of all possible phase screens)

$$\langle \sigma_A \sigma_B \rangle = \langle e^{i(\phi_A + \phi_B)} \rangle \tag{9.1}$$

it is necessary to know the probability that a path segment starting with $x = x_A$ and finishing with $x = x_B$ passes through an odd number of these lines (producing a sign change). The lines are randomly distributed with mean spacing $\bar{a}$. A Poisson distribution [7] with an average of $\frac{1}{\bar{a}}$ lines per unit interval of x is appropriate, so the probability $P(n)$ that there are n lines in an interval of length $|\Delta x| = |x_B - x_A|$ is

$$P(n) = \frac{|\Delta|^n e^{-|\Delta|}}{n!} \tag{9.2}$$

where $\Delta = \frac{\Delta x}{\bar{a}}$

so

$$\begin{aligned} P(n\ is\ odd) &= e^{-|\Delta|} \sum_{i=0}^{\infty} \frac{|\Delta|^{2i+1}}{(2i+1)!} \\ &= e^{-|\Delta|} \sinh |\Delta| \end{aligned} \tag{9.3}$$

$$\begin{aligned} P(n\ is\ even) &= e^{-|\Delta|} \sum_{i=0}^{\infty} \frac{|\Delta|^{2i}}{(2i)!} \\ &= e^{-|\Delta|} \cosh |\Delta| \end{aligned} \tag{9.4}$$

$$\begin{aligned} \langle \sigma_A \sigma_B \rangle &= (+1) \times P(n\ is\ even) + (-1) \times P(n\ is\ odd) \\ &= e^{-|\Delta|}(\cosh |\Delta| - \sinh |\Delta|) \\ &= e^{-2|\Delta|} \end{aligned} \tag{9.5}$$

So whereas for the periodic model the exponential form was an approximation, it is an exact result for the random phase screens model. Similarly, the result for the averaged sign for K^*K is now exact, i.e.

$$\langle e^{i(\phi''_A + \phi''_B - \phi'_A - \phi'_B)} \rangle = e^{-\frac{2}{\bar{a}}T} \tag{9.6}$$

where T is the "total non-overlapping length" ($T > 0$).

Since a line in the region where two path segments "overlap" produces a π phase change

for both paths (which has no net effect), this region may be treated in the same way as a region where there are no paths.

9.2.3 Investigation of $\langle K \rangle$

Having an exact result means that there are no neglected terms, which in turn allows a slightly more detailed investigation to be attempted. From the previous chapter it is expected that $\langle K \rangle$ will be 0 for finite propagation when $\langle \sigma_A \sigma_B \rangle = e^{-2\Delta}$. It is of interest to build $\langle K \rangle$ for a finite z-interval (or time interval) by composition of infinitesimal steps (in the z-direction). The result for multiple steps is obtained by repeated convolution of the function

$$\begin{aligned} K_1 &= \sqrt{\frac{\nu}{\pi i}}\, e^{i\nu\Delta x^2 - \frac{2}{\bar{a}}|\Delta x|} \\ &= \sqrt{\frac{\rho}{\pi i \bar{a}^2}} e^{i\rho\Delta^2 - 2|\Delta|} \end{aligned} \tag{9.7}$$

where

$$\rho = \nu \bar{a}^2 \tag{9.8}$$

The Fourier transform of the exponential is:

$$I(k,\rho) = I_+ + I_- \tag{9.9}$$

where k is the "Fourier variable" (in this section) and

$$I_\pm = \int_{y=0}^{\infty} e^{(i\rho y^2 - 2y \pm iky)} dy \tag{9.10}$$

The complex conjugate $I_\pm^*$ can be written in terms of an integral of the type considered in [5] i.e.

$$I_\pm^* = i\sqrt{i}\, e^{-\rho w_0^2} J \tag{9.11}$$

with

$$J = \int_{w=-i\infty}^{w_0} e^{\rho w^2} dw \tag{9.12}$$

where

$$w_0 = \sqrt{i}\left(-\frac{i}{\rho} \mp \frac{k}{2\rho}\right) \tag{9.13}$$

This is useful because ρ is proportional to N and is therefore a large parameter when many steps are composed. Specifically, this is true when considering a finite interval in the z-direction which is divided into a large number, N, of subintervals. For a given finite

interval of z, the length of a subinterval (which is written as $c\epsilon$ in the definition (6.3) for ν) is inversely proportional to N ($N \to \infty$ being the "continuum limit" of infinitesimal step size). So ν and hence ρ (equation (9.8)) are proportional to N. Using the result for the leading terms in the asymptotic expansion from [5]

$$J \sim (2\rho w_0)^{-1} \exp(\rho w_0^2) + iS(w_0, \rho) \left(\frac{\pi}{\rho}\right)^{\frac{1}{2}} \qquad as \ \rho \to \infty \tag{9.14}$$

where $S(w_0, \rho)$ is the Stokes multiplier [5]

gives, for the limiting form of the Fourier transform:

$$I \sim \frac{-i}{\left(\frac{k}{2}\right)^2 + 1} - \sqrt{\frac{i\pi}{\rho}} e^{\frac{i}{\rho}\left(\left(\frac{k}{2}\right)^2 - 1\right)} e^{-\frac{|k|}{\rho}} \qquad as \ \rho \to \infty \tag{9.15}$$

for $|k| \geq 2$

In fact this result does not hold for $|k| < \sqrt{\rho}$ (hence the $|k| < 2$ case is not given)

For $|k| < \sqrt{\rho}$ the standard result

$$\int_{t=z}^{\infty} e^{-t^2} dt = \frac{\sqrt{\pi}}{2} - e^{-z^2} \left(z + \frac{2}{3} z^3 + \ldots\right) \tag{9.16}$$

may be used in the expression

$$I_{\pm} = \sqrt{\frac{i}{\rho}} e^{z^2} \int_{t=z}^{\infty} e^{-t^2} dt \tag{9.17}$$

where $z = \sqrt{\frac{i}{\rho}} \left(1 \mp \frac{ik}{2}\right)$

to obtain the behaviour of $I_{\pm}$ in the $\rho \to \infty$ limit. Substituting this into equation (9.9) and expanding the exponentials in power series (permissible since the parameter $\rho \gg 1$) allows a factor $\left(1 - C\sqrt{\frac{i}{\rho}}\right)$ to be extracted, as in section 8.4.4 (where C is a positive constant independent of the "Fourier variable", k). Composition of N stages is achieved by taking the N^{th} power (of K_1) and then the inverse Fourier transform. So the contribution for $|k| < \sqrt{\rho}$ tends to zero as N becomes large, as before. Specifically, one expects

$$\left(1 - \frac{B}{\sqrt{N}}\right)^N \sim e^{-B\sqrt{N}} \qquad as \ N \to \infty \tag{9.18}$$

(for a quantity B which is not a function of N and satisfies $\Re(B) > 0$)

whereas in the usual case one has the standard result

$$\left(1 - \frac{B}{N}\right)^N \to e^{-B} \qquad as \ N \to \infty \tag{9.19}$$

For $|k| > \sqrt{\rho}$ the final term in the expression (9.15) for I will be dominant, so the function of k to be integrated is

$$f(k) = e^{\frac{iN}{\rho}\left(\left(\frac{k}{2}\right)^2 - 1\right)} e^{-\frac{N|k|}{\rho}} e^{-ikY} \tag{9.20}$$

where Y is proportional to the total displacement for N stages, $x_{final} - x_{initial}$
The function $f(k)$ is independent of N and ρ because $\frac{N}{\rho}$ is a (positive) constant (R say).
The expression to be evaluated is

$$\int_{\sqrt{\rho}}^{\infty} f(k)dk + \int_{-\infty}^{-\sqrt{\rho}} f(k)dk = \int_{\sqrt{\rho}}^{\infty} (f_+(k) + f_-(k))\, dk \tag{9.21}$$

where

$$f_\pm(k) = e^{iR\left(\left(\frac{k}{2}\right)^2 - 1\right) - (R \pm iY)k} \tag{9.22}$$

(and $R = \frac{N}{\rho}$)
calculation (i.e. developing an asymptotic series by repeated integration by parts in the standard way [4]) shows that

$$\left| \int_{\sqrt{\rho}}^{\infty} f_\pm(k)dk \right| \sim \frac{1}{R\sqrt{\rho}} e^{-R\sqrt{\rho}} \qquad as \;\; \rho \to \infty \tag{9.23}$$

so these contributions are indeed negligible as $\rho \to \infty$. Having verified that in general $\langle K \rangle$ is zero for propagation over a finite time interval, it is desirable to check that probability is conserved under "propagation" by $\langle K^*K \rangle$.

9.2.4 Conservation of probability

This calculation differs from that performed for *periodic* phase screens (section 8.4.3) because it is now necessary to introduce averaging. This requires the result (for $z_B > z_A$)

$$\begin{aligned} &\langle \psi^*(\underline{r}'_B, z_B)\psi(\underline{r}''_B, z_B) \rangle \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \langle K^*(\underline{r}'_B, z_B; \underline{r}'_A, z_A) K(\underline{r}''_B, z_B; \underline{r}''_A, z_A) \rangle \psi^*(\underline{r}'_A, z_A)\psi(\underline{r}''_A, z_A)\, d^2\underline{r}'_A\, d^2\underline{r}''_A \end{aligned} \tag{9.24}$$

which holds because the choice of initial wavefunction is not affected by averaging. Setting $\underline{r}''_B = \underline{r}'_B$ and integrating over this final position:

$$\begin{aligned} &\int_{-\infty}^{\infty} \langle \psi^*(\underline{r}_B, z_B)\psi(\underline{r}_B, z_B) \rangle d^2\underline{r}_B \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \langle K^*(\underline{r}_B, z_B; \underline{r}'_A, z_A) K(\underline{r}_B, z_B; \underline{r}''_A, z_A) \rangle \psi^*(\underline{r}'_A, z_A)\psi(\underline{r}''_A, z_A) d^2\underline{r}'_A d^2\underline{r}''_A d^2\underline{r}_B \end{aligned} \tag{9.25}$$

The "averaged sign" for K^*K in the special case $\underline{r}''_B = \underline{r}'_B$ is required. The general result is $e^{-\frac{2}{a}T}$. When $\underline{r}''_B = \underline{r}'_B$ the total non-overlapping length T is $|x''_A - x'_A|$. The dependence on $\underline{r}_B$ is the same as before (section 8.4.3), so the integral over $\underline{r}_B$ is the same, giving

$$\int_{-\infty}^{\infty} \langle K(\underline{r}_B, z_B; \underline{r}''_A, z_A) K^*(\underline{r}_B, z_B; \underline{r}'_A, z_A) \rangle d^2\underline{r}_B = F_r(\underline{r}''_A, \underline{r}'_A)\delta^2(\underline{r}'_A - \underline{r}''_A) \tag{9.26}$$

The function F used in the periodic case (defined by equation (8.22)) is replaced by a new function F_r in the random case which has

$$\exp\left(i\pi\left[\frac{x''_A-\alpha}{a}\right]+i\pi\left[\frac{x'_A-\alpha}{a}\right]\right)$$

replaced by

$$\exp\left(-\frac{2}{\bar{a}}|x''_A-x'_A|\right)$$

However, the only property (of F_r) required for the conservation of probability is that

$$F_r(\underline{r}_A,\underline{r}_A)=1$$

and this is indeed true, so that

$$\int_{-\infty}^{\infty}\langle\psi^*(\underline{r}_B,z_B)\psi(\underline{r}_B,z_B)\rangle d^2\underline{r}_B=\int_{-\infty}^{\infty}\psi^*(\underline{r}_A,z_A)\psi(\underline{r}_A,z_A)d^2\underline{r}_A \qquad (9.27)$$

Again, this process may be iterated to show that probability is conserved for finite time intervals.

9.2.5 Types of path

It is instructive to consider the types of path (pairs) which are likely to make an important contribution to $\langle K^*K\rangle$. The presence of the factors $\exp\left(-\frac{2}{\bar{a}}T\right)$ means that such paths are those for which the non-overlapping length T is small compared to $\bar{a}$ (the mean width of a phase strip) for the majority of the stages from which the path is composed. There are three types of stage for which T is zero, they are:

Type 1 Coincident path segments (of any slope)

Type 2 Crossing path segments with slopes of equal magnitude

Type 3 Parallel zero slope path segments

These are shown in figure 9.3. In order to apply the constraint successfully, it is expected that paths with low slope, i.e. $\frac{|\Delta x|}{\delta z}\ll 1$ (for all stages) should be "preferred" compared to those of large slope ($\frac{|\Delta x|}{\delta z}\gg 1$). Consequently, it is reasonable that path pairs constructed predominantly from "type 3" segments should contribute strongly to $\langle K^*K\rangle$. However, the slope of the path segments in "type 1" and "type 2" stages can be of any size. The fact that these types of stage are likely to contribute strongly to $\langle K^*K\rangle$ is an artefact of the phase screens model. Consequently, it seems advisable to return to the "mirror planes" method of applying the constraint.

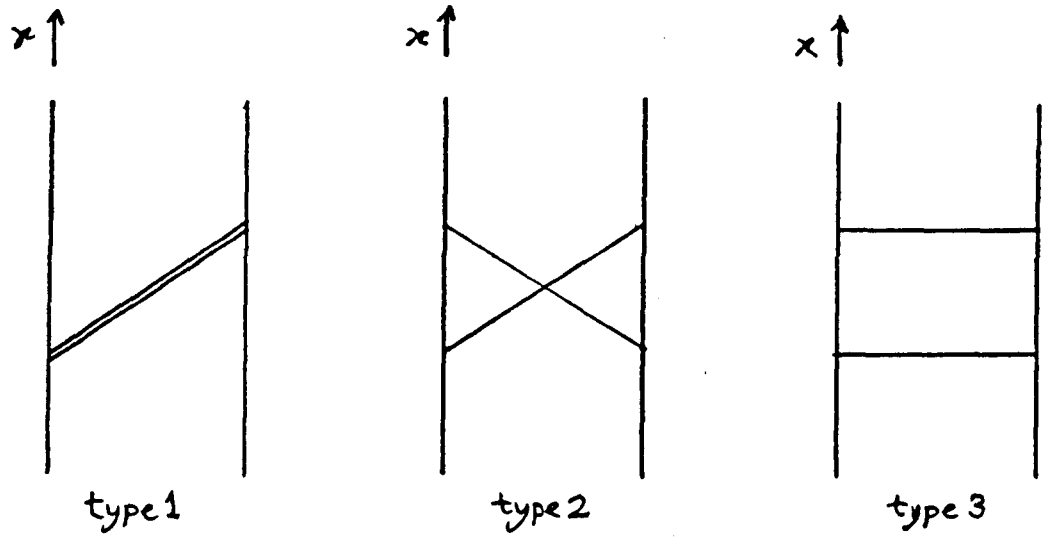


Figure 9.3: Special cases for a single 1D stage

9.2.6 Summary

In chapter 8 some results are obtained for a version of the Phase screens model by making heuristic approximations. In this section (i.e. section 9.2) similar expressions are obtained as exact results for a modified version of the same model. This increases confidence in the results. From a different point of view, it suggests that the modification made to the model is a beneficial one.

In this section, use has been made of the "phase counting planes " construction. This is merely a change of emphasis which is beneficial for the type of calculation performed in this section. The same idea could have been used in chapter 8 but was omitted as an unnecessary distraction from the "physical" aspects of the model.

The behaviour of $\langle K \rangle$ and $\langle K^*K \rangle$ is the same as in chapter 8. Although the general behaviour of $\langle K^*K \rangle$ is correct, consideration of the details reveals undesirable features of the phase screens model. These undesirable features are a consequence of the way the constraint is (approximately) enforced in the phase screens model. It seems sensible to apply the "beneficial modification" of this section (i.e. the introduction of "randomness") to the "mirror planes" model, which, as an exact way of enforcing the constraints, shouldn't suffer from the undesirable features noticed in the phase screens case.

9.3 Mirror planes

9.3.1 Introduction

Returning to the original ("direct") approach to the "mirror planes" model (i.e. equation (6.1)), it is hoped that replacing the regularly spaced stack of strips ("mirror planes") with a randomly spaced version will alleviate some of the problems previously encountered (in section 6.3).

9.3.2 Preliminaries

It is expected that the behaviour of the expression for the (single stage) one-dimensional propagator (equation (6.1)) can be improved if it is averaged over the "shifts" β and the separation between mirror planes ("lane width"). Averaging over β alone is not sufficient because the first sum is independent of the position of the planes —provided that there are no planes between the specified initial and final points (in which case $K = 0$). Averaging over "shifts" (only) was introduced in section 8.4 for the *Phase screens* model. It is desired to go further and include "randomness" — as carried out for Phase screens in section 9.2. The introduction of "randomness" will give a "weighted" average over "lane width".

9.3.3 Single stage propagation

To formulate an expression for the averaged single stage propagator, it is necessary to consider the consequences of having a randomly spaced stack of constraint strips (referred to as "planes" here).

If "planes" are distributed at random, then the probability distribution for the separation, a, of the "planes" (with average pair spacing $\bar{a}$) is proportional to $\exp(-\frac{a}{\bar{a}})$. The initial position of the particle is a point (the "initial point") which lies in the gap between a pair of planes. This pair of planes is then (rigidly) shifted relative to the particle in such a way that all shifts consistent with the "initial point" remaining inside the lane are applied (i.e. a total distance of a between extreme shifts). If the specified final point lies outside the lane defined by the pair of planes then K is zero, otherwise it may be calculated using image charges . This means that there is no contribution when the (1D) separation between the initial and final points is greater than the "lane width", i.e. $|\Delta x| > a$. Applying this prescription for the calculation of the averaged single stage one-dimensional propagator, $\langle K_1 \rangle$:

$$\langle K_1 \rangle = \int_{a=|\Delta x|}^{\infty} da\, p(a) \int_{\beta=\beta_1}^{\beta_2} \frac{1}{a} K(\Delta x, \beta, a) d\beta \tag{9.28}$$

where

$$p(a) = \frac{1}{\bar{a}} e^{-\frac{a}{\bar{a}}}$$

$p(a)da$ is the probability of finding two adjacent planes a distance a apart when the average spacing is $\bar{a}$ and the planes are randomly distributed [14]. The probability distribution for the shifts β is uniform. The "shift" β is defined to be the (initial) distance from the particle to the mirror plane "below" it, i.e. the nearest plane in the direction of x decreasing. The cases $\Delta x \geq 0$ and $\Delta x < 0$ must be considered separately:

- for $\Delta x \geq 0 : \beta_1 = 0,\ \beta_2 = a - |\Delta x|$
- for $\Delta x < 0 : \beta_1 = |\Delta x|,\ \beta_2 = a$

So substituting for $K(\Delta x, \beta, a)$ from equation (6.1) gives

$$\begin{aligned}
&\langle K_1(\Delta x; \bar{a}, \nu)\rangle \\
&= \sqrt{\frac{\nu}{\pi i}} \frac{1}{\bar{a}} \int_{a=|\Delta x|}^{\infty} da\, e^{-\frac{a}{\bar{a}}} \sum_{n=-\infty}^{\infty} \left[\frac{a - |\Delta x|}{a} e^{i\nu(\Delta x + 2an)^2} - \frac{1}{a} \int_{\beta=\beta_1}^{\beta_2} e^{i\nu(\Delta x + 2\beta + 2an)^2} d\beta \right]
\end{aligned} \tag{9.29}$$

It is possible to use the formula

$$\sum_{n=-\infty}^{\infty} \int_{\beta=|\Delta x|}^{a} e^{i\nu(y+2\beta+2an)^2} d\beta = \sum_{N=-\infty}^{\infty} \int_{B=0}^{(a-|\Delta x|)} e^{i\nu(-y+2B+2aN)^2} dB \tag{9.30}$$

where $N = -(n+1)$

to rewrite equation (9.29) for $\Delta x < 0$ (i.e. set $y = \Delta x$) so that there is no explicit dependence on the sign of Δx, i.e.

$$\begin{aligned}
&\langle K_1(\Delta x; \bar{a}, \nu)\rangle \\
&= \sqrt{\frac{\nu}{\pi i}} \int_{a=|\Delta x|}^{\infty} \frac{1}{\bar{a}} da\, e^{-\frac{a}{\bar{a}}} \\
&\quad \times \left[\left(1 - \frac{|\Delta x|}{a}\right) \sum_{n=-\infty}^{\infty} e^{i\nu(|\Delta x| + 2an)^2} - \frac{1}{a} \sum_{N=-\infty}^{\infty} \int_{\beta=0}^{(a-|\Delta x|)} e^{i\nu(|\Delta x| + 2\beta + 2aN)^2} d\beta \right]
\end{aligned} \tag{9.31}$$

where

$N = n$, $n = n$ for $\Delta x \geq 0$,

$N = -(n+1)$, $n = -n$ for $\Delta x < 0$.

The distinction between the two cases is not important for the infinite summations but is noted here because the way corresponding terms are "paired" if the sums are combined may be significant for numerical evaluation of a truncated version.

Introducing dimensionless parameters and dimensionless variables of integration $A = \frac{a}{\bar{a}}$, $\phi = \frac{\beta}{a-|\Delta x|}$ (and writing the average of K_1 as $\bar{K}_1$ when dimensionless quantities are used) equation (9.31) becomes

$$\begin{aligned} \bar{K}_1 &= \frac{1}{\bar{a}}\sqrt{\frac{\rho}{\pi i}}\int_{A=|\Delta|}^{\infty} dA\, e^{-A}\left(1-\frac{|\Delta|}{A}\right) \\ &\times\left[\sum_{n=-\infty}^{\infty} e^{i\rho(|\Delta|+2An)^2} - \sum_{N=-\infty}^{\infty}\int_{\phi=0}^{1} e^{i\rho(|\Delta|+2\phi(A-|\Delta|)+2AN)^2} d\phi\right] \end{aligned} \tag{9.32}$$

where $\rho = \nu\bar{a}^2$, $|\Delta| = \frac{|\Delta x|}{\bar{a}}$

Another version of the result (9.32) may be obtained by using (a special case of) the Poisson summation formula [28] (also appendix F)

i.e.

$$\sum_{n=-\infty}^{\infty} f(n) = \sum_{m=-\infty}^{\infty}\int_{-\infty}^{\infty} f(x) e^{i(2\pi m)x} dx \tag{9.33}$$

to rewrite the sum

i.e.

$$\begin{aligned} \bar{K}_p &= \frac{1}{2\bar{a}}\int_{A=|\Delta|}^{\infty} dA\, e^{-A}\left(1-\frac{|\Delta|}{A}\right) \\ &\times\left[\sum_{m=-\infty}^{\infty}\frac{1}{A} e^{-\frac{i\pi m}{A}\left(|\Delta|+\frac{\pi m}{4\rho A}\right)} - \sum_{m=-\infty}^{\infty}\frac{1}{A}\int_{\phi=0}^{1} d\phi\, e^{-\frac{i\pi m}{A}\left(|\Delta|+2\phi(A-|\Delta|)+\frac{\pi m}{4\rho A}\right)}\right] \end{aligned} \tag{9.34}$$

The notation "$\bar{K}_p$" is introduced to distinguish the "Poisson transformed" version of the result from the "original". When the formula is written in this "$\bar{K}_p$" form, the integral over ϕ may be evaluated explicitly in terms of elementary functions, i.e.

$$\begin{aligned} g(q) &\equiv \int_{\phi=0}^{1} e^{-iq\phi} d\phi \\ &= \frac{i}{q}\left(e^{-iq}-1\right) \end{aligned} \tag{9.35}$$

where $q = 2\pi m\left(1-\frac{|\Delta|}{A}\right)$

Similar approaches are useful for the evaluation of sums in problems from solid state physics.

For the "Poisson transformed" expression (i.e. for $\bar{K}_p$) there is one term in each sum which has zero exponent, these terms should be matched (this correspondence is enough to determine all pairings). In fact, these terms are the same and hence cancel out when the sums are combined (i.e. $g(0) = 1$ so the $m = 0$ term in equation (9.34) vanishes).

9.3.4 A simple case

When $\Delta = 0$ the result (9.32) simplifies to

$$\bar{K}_1(0) = \frac{1}{\bar{a}}\sqrt{\frac{\rho}{\pi i}}\int_{A=0}^{\infty} dA\, e^{-A}\left[\sum_{n=-\infty}^{\infty} e^{4i\rho n^2 A^2} - \sum_{N=-\infty}^{\infty}\int_{\phi=0}^{1} e^{4i\rho(\phi+N)^2 A^2} d\phi\right] \tag{9.36}$$

where $\bar{K}_1(0)$ denotes $\bar{K}_1$ evaluated at $\Delta = 0$.

This statement may be justified by considering the limit of $\bar{K}_1$ as $|\Delta|$ tends to zero through positive values. For the $n = 0$ term this involves verifying that

$$\lim_{|\Delta|\to 0+} \frac{1}{\bar{a}}\sqrt{\frac{\rho}{\pi i}} e^{i\rho|\Delta|^2}\int_{A=|\Delta|}^{\infty} e^{-A}\left(1 - \frac{|\Delta|}{A}\right) dA = \frac{1}{\bar{a}}\sqrt{\frac{\rho}{\pi i}} \tag{9.37}$$

This is indeed true since (from the definition of $E_1(x)$ [1, 21])

$$|\Delta|\int_{A=|\Delta|}^{\infty} \frac{e^{-A}}{A} dA = |\Delta| E_1(|\Delta|) \tag{9.38}$$

and the standard result [1, 21]

$$E_1(x) = -\gamma - \ln x + \int_0^x \frac{1 - e^{-u}}{u} du \tag{9.39}$$

(where γ is Euler's constant)

for the Exponential integral $E_1(x)$ shows that

$$\lim_{|\Delta|\to 0+} |\Delta| E_1(|\Delta|) = 0 \tag{9.40}$$

since $x \ln x \to 0$ as $x \to 0+$.

For the other terms in the sum over n, the only difference is the presence of an "oscillating factor" $\exp(i\rho(4|\Delta|An + 4A^2n^2))$ in the integral over A. This factor has unit modulus so it is expected that

$$\lim_{|\Delta|\to 0+} |\Delta|\int_{A=|\Delta|}^{\infty} \frac{\left(e^{-A}\right)\left(e^{i\rho(|\Delta|+2An)^2}\right)}{A} dA = 0 \tag{9.41}$$

holds for $n \neq 0$ as well as $n = 0$.

Similarily, for the terms in the sum over N, the introduction of ϕ into the exponent of the "oscillating factor" is not expected to cause problems. Indeed, none of these results is surprising if the integrands are sketched over the integral of integration.

Perhaps the simplest way to obtain the $|\Delta| = 0$ case of the Poisson transformed formula, $\bar{K}_p$, is to apply the Poisson transformation to the $|\Delta| = 0$ expression (9.36), the result is

$$\bar{K}_p(0) = \frac{1}{\bar{a}}\sum_{m=1}^{\infty}\int_{A=0}^{\infty} dA \frac{1}{A} e^{-A - \frac{1}{4\rho}\left(\frac{\pi m}{A}\right)^2} \tag{9.42}$$

where $\bar{K}_p(0)$ denotes $\bar{K}_p$ evaluated at $\Delta = 0$

However, it is also possible to deduce that taking the limit $|\Delta| \to 0+$ in the general expression (9.34) gives the same result. The parts of the terms in the sums of the form

$$\int_{A=|\Delta|}^{\infty} dA\, e^{-A} \left(\frac{|\Delta|}{A}\right) \frac{1}{A} e^{-\frac{i\pi m}{A}\left(|\Delta|+2\phi(A-|\Delta|)+\frac{\pi m}{4\rho A}\right)}$$

(where ϕ may be zero)

will be smaller in magnitude than $E_2(|\Delta|)$ due to the oscillating (unit modulus) exponential factor and hence will tend to zero as $|\Delta| \to 0+$. This is plausible since the smaller (in magnitude) the values of A, the more rapid the oscillation and applying integration by parts to

$$\begin{aligned} E_2(z) &\equiv \int_1^{\infty} \frac{e^{-zt}}{t^2} dt \qquad \Re z > 0 \\ &= z \int_z^{\infty} \frac{e^{-u}}{u^2} du \end{aligned} \tag{9.43}$$

gives

$$E_2(z) = e^{-z} - zE_1(z) \tag{9.44}$$

and hence (using equation (9.40))

$$\lim_{|\Delta| \to 0+} E_2(|\Delta|) = 1 \tag{9.45}$$

i.e. the "marginal" case.

9.3.5 Asymptotics

The asymptotic expansions to be considered are those for $\rho \to \infty$ and $\rho \to 0$. The parameter ρ is $\nu\bar{a}^2$ (ν is defined by equation (6.2) or equation (6.3)) so these limits are related to those considered earlier (in section 6.2 and section 7.2) and discussed in terms of vakonomic and nonholonomic regimes.

Beginning with the special case $\Delta = 0$ and taking the limit $\rho \to \infty$ in the expression (9.36) for $\bar{K}_1$ (the limit $\rho \to 0$ for this form of the infinitesimal propagator is not instructive) gives,

$$\bar{a}\sqrt{\frac{\pi i}{\rho}}\, \bar{K}_1 = 1 + O\left(\frac{\ln \rho}{\sqrt{\rho}}\right) \text{ as } \rho \to \infty \tag{9.46}$$

(where use has been made of the "order symbol" O [35])

The leading term comes from the $n = 0$ term of the first sum (the "zero bounce term").

For this term the integration over A is particularly straightforward,

i.e.

$$\int_{A=0}^{\infty} dA\, e^{-A} = 1 \tag{9.47}$$

The other terms in the first sum are of order $\frac{1}{\sqrt{\rho}}$ as $\rho \to \infty$. This is shown by the asymptotic expansion of the relevant integral, i.e.

$$\int_{A=0}^{\infty} e^{-A+4i\rho n^2 A^2} dA = \frac{1}{4}\sqrt{\frac{\pi i}{\rho n^2}} e^{\frac{i}{16\rho n^2}} + O\left(\frac{1}{\rho n^2}\right) \quad \text{as } \rho \to \infty \tag{9.48}$$

To avoid divergence of the individual sums in equation (9.36) they are combined into a single summation. It should be noted that if the integral in equation (9.48) is written in terms of a standard error function, then its argument depends upon $\frac{1}{\sqrt{\rho n^2}}$ which tends to zero as $\rho \to \infty$.

The terms of the second sum in equation (9.36), excluding $N = 0$ and $N = -1$, are also of order $\frac{1}{\sqrt{\rho}}$ as $\rho \to \infty$, each term taking the form

$$\int_{A=0}^{\infty} dA\, e^{-A} \int_{\phi=0}^{1} e^{4i\rho(\phi+N)^2 A^2} d\phi = \int_{\phi=0}^{1} d\phi \left[\frac{\sqrt{\pi i}}{4\sqrt{\rho(N+\phi)^2}} e^{\frac{i}{16\rho(N+\phi)^2}} + O\left(\frac{1}{\rho(N+\phi)^2}\right)\right] \quad \text{as } \rho \to \infty \tag{9.49}$$

The $N = 0$ and $N = -1$ terms require separate consideration since $N + \phi = 0$ for $N = 0$, $\phi = 0$ and $N = -1$, $\phi = 1$. It is convenient to write these terms in a slightly different form, obtained by substituting $A = \frac{\alpha}{a}$ and $B = \frac{\beta}{a}$ in equation (9.31), and setting $\Delta = 0$. The resulting expression is, for the $N = 0$ term

$$T_0 = \int_{A=0}^{\infty} dA \frac{1}{A} e^{-A} \int_{B=0}^{A} e^{4i\rho B^2} dB \tag{9.50}$$

In fact, the $N = -1$ term can also be written in this form if B is replaced by B' using $B' = A - B$. The substitutions $A = \frac{\alpha}{\sqrt{\rho}}$ and $B = \frac{u}{\sqrt{\rho}}$ give

$$\begin{aligned} T_0 &= \frac{1}{\sqrt{\rho}} \int_{\alpha=0}^{\infty} d\alpha \frac{1}{\alpha} e^{-\frac{\alpha}{\sqrt{\rho}}} \int_{u=0}^{\alpha} e^{4iu^2} du \\ &= T_a + T_b \end{aligned} \tag{9.51}$$

where

$$T_a = \frac{1}{\sqrt{\rho}} \int_{\alpha=0}^{1} d\alpha \frac{1}{\alpha} e^{-\frac{\alpha}{\sqrt{\rho}}} \int_{u=0}^{\alpha} e^{4iu^2} du \tag{9.52}$$

$$T_b = \frac{1}{\sqrt{\rho}} \int_{\alpha=1}^{\infty} d\alpha \frac{1}{\alpha} e^{-\frac{\alpha}{\sqrt{\rho}}} \int_{u=0}^{\alpha} e^{4iu^2} du \tag{9.53}$$

In expression (9.53) rapid oscillation of the integrand means that it is a reasonable approximation to replace the upper limit of integration by ∞, for the integral over u. So

$$\begin{aligned} T_b &\approx \frac{1}{\sqrt{\rho}} \int_{A=\frac{1}{\sqrt{\rho}}}^{\infty} dA \frac{1}{A} e^{-A} \int_{u=0}^{\infty} e^{4iu^2} du \\ &= \frac{1}{4}\sqrt{\frac{\pi i}{\rho}} E_1\left(\frac{1}{\sqrt{\rho}}\right) \end{aligned} \tag{9.54}$$

So, using the standard [1, 21] expansion (obtained from (9.39) by expanding the integral as a power series) for the exponential integral $E_1(x)$

$$T_b = \frac{\sqrt{\pi i}}{4}\left(\frac{1}{2}\frac{\ln\rho}{\sqrt{\rho}} - \frac{\gamma}{\sqrt{\rho}} + \frac{1}{\rho} + O(\rho^{-\frac{3}{2}})\right) \text{ as } \rho \to \infty \tag{9.55}$$

where γ is Euler's constant.

Considering now expression (9.52)

$$\Re(T_a) = \frac{1}{\sqrt{\rho}} \int_{\alpha=0}^{1} d\alpha \frac{1}{\alpha} e^{-\frac{\alpha}{\sqrt{\rho}}} \int_{u=0}^{\alpha} \cos(4u^2) du \tag{9.56}$$

$$\Im(T_a) = \frac{1}{\sqrt{\rho}} \int_{\alpha=0}^{1} d\alpha \frac{1}{\alpha} e^{-\frac{\alpha}{\sqrt{\rho}}} \int_{u=0}^{\alpha} \sin(4u^2) du \tag{9.57}$$

so

$$|\Re(T_a)| < \frac{1}{\sqrt{\rho}} \int_{\alpha=0}^{1} d\alpha \frac{1}{\alpha} e^{-\frac{\alpha}{\sqrt{\rho}}} \int_{u=0}^{\alpha} 1 du \tag{9.58}$$

$$|\Im(T_a)| < \frac{1}{\sqrt{\rho}} \int_{\alpha=0}^{1} d\alpha \frac{1}{\alpha} e^{-\frac{\alpha}{\sqrt{\rho}}} \int_{u=0}^{\alpha} 1 du \tag{9.59}$$

hence

$$\begin{aligned} |T_a| &< \sqrt{2}\frac{1}{\sqrt{\rho}} \int_{\alpha=0}^{1} d\alpha \frac{1}{\alpha} e^{-\frac{\alpha}{\sqrt{\rho}}} \int_{u=0}^{\alpha} 1 du \\ &= \frac{\sqrt{2}}{\sqrt{\rho}} \int_{\alpha=0}^{1} d\alpha\, e^{-\frac{\alpha}{\sqrt{\rho}}} \\ &= \sqrt{2}(1 - e^{-\frac{1}{\sqrt{\rho}}}) \\ &= \frac{\sqrt{2}}{\sqrt{\rho}} + O\left(\frac{1}{\rho}\right) \text{ as } \rho \to \infty \end{aligned} \tag{9.60}$$

so T_b is the dominant contribution as $\rho \to \infty$, which gives the required result (9.46).

In order to take the limit $\rho \to 0$, the "transformed" version of the infinitesimal propagator, $\bar{K}_p$, is required, i.e. (for the $\Delta = 0$ case) expression (9.42). There is a large parameter $\lambda = \frac{(\pi m)^2}{4\rho}$ (so $\lambda \to \infty$ as $\rho \to 0$). Standard saddle point methods [26] may be applied. A substitution $A = e^B$ may be made to clarify the situation. It is expected that, of the three saddle points, only the one nearest the path of integration will contribute, hence

$$\bar{K}_p(\Delta = 0) \sim \frac{1}{\bar{a}} \sum_{m=1}^{\infty} \sqrt{-\frac{2}{3}\left(\frac{2\pi\rho}{im^2}\right)^{\frac{1}{3}}} e^{-\frac{3}{2}\left(\frac{i\pi^2 m^2}{2\rho}\right)^{\frac{1}{3}}} \text{ as } \rho \to 0 \tag{9.61}$$

So, as well as decreasing as ρ tends to zero, the terms of the series decrease as m^2 increases in size.

9.3.6 Further asymptotics

Having considered the $\Delta = 0$ case in both the limits $\rho \to \infty$ and $\rho \to 0$ it is desirable to investigate the general case by considering how things differ when Δ is non-zero.

For $\Delta \neq 0$ the functional forms of the expressions are more complicated. However, for $\rho \to \infty$, $\Delta \neq 0$ means that $N + \phi = 0$ is no longer a problem. Such cases no longer need to be treated separately. It is expected that the "zero bounce" term again dominates and that the magnitudes of the other terms relative to this now tend to zero like $\frac{1}{\sqrt{\rho}}$ or faster as $\rho \to \infty$, i.e.

$$\bar{a}\sqrt{\frac{\pi i}{\rho}}\bar{K}_1 = e^{i\rho|\Delta|^2}\int_{A=|\Delta|}^{\infty} e^{-A}\left(1 - \frac{|\Delta|}{A}\right)dA + O\left(\frac{1}{\sqrt{\rho}}\right) \text{ as } \rho \to \infty \qquad (9.62)$$

upon evaluating the integral this becomes

$$e^{-|\Delta|+i\rho|\Delta|^2} - e^{i\rho|\Delta|^2}|\Delta|E_1(|\Delta|)$$

the first term is the random phase screens result which is recovered in the limit $|\Delta| \to 0+$.

For the limit $\rho \to 0$, making the substitution $A = \frac{\alpha}{\sqrt{\rho}}$ in the formula (9.34) for $\bar{K}_1$ produces an expression in which Δ appears only in the combination $\sqrt{\rho}|\Delta|$. Consequently, the Δ dependence is expected to become less prominent as $\rho \to 0$ since its "scale length" is $\frac{1}{\sqrt{\rho}}$.

To interpret the limiting expressions it is necessary to use the fact that the parameter ρ is proportional to $\frac{\bar{a}^2}{\delta z}$. As ρ tends to infinity (which requires $\delta z \to 0$ since it is desired that $\bar{a} \to 0$), the mean spacing of the mirror planes strips (i.e. $\bar{a}$) becomes large compared to the "length" of a single stage (δz) in the z-direction ("wide lanes"). Similarly, as ρ tends to zero, the average spacing between "mirror planes" becomes comparatively small ("narrow lanes"). The result that the "zero bounce" path dominates as $\rho \to \infty$ is reasonable since in this case the "bouncing path segments" have, on average, gradients much larger in size than $\frac{|\Delta x|}{\delta z}$ (the magnitude of the gradient of the direct path).

For path segments of given gradient (i.e. $\frac{|\Delta x|}{\delta z} = constant$), $\rho \to \infty$ (with $\bar{a} \to 0$) means that $\Delta \to 0$ (i.e. $(\frac{|\Delta x|}{\delta z}/\frac{\bar{a}^2}{\delta z}) \to 0$ which is $\frac{|\Delta x|}{\bar{a}} \to 0+$) and in this limit the random phase screens result is recovered. It is more relevant for quantum mechanics that this result holds for path segments with $|\Delta x| \sim \delta z^{\frac{1}{2}}$ as $\delta z \to 0$. In this case $\frac{|\Delta x|}{\delta z^{\frac{1}{2}}} \sim 1$ as $\delta z \to 0$ so $\frac{|\Delta x|}{\delta z^{\frac{1}{2}}}/\sqrt{\rho} \to 0$ as $\rho \to \infty$ and hence $|\Delta| \to 0+$ as $\rho \to \infty$ which means that the phase

screens result is again recovered. These results are for the "path segments" form of the single stage propagator. For the transformed version $\bar{K}_p$, in the special case $\Delta = 0$, the result (9.42) is similar to a sum over modes, with a type of averaging over the lane widths. When $\rho \to 0$ it is clear that the largest contribution to the sum comes from the lowest modes: these correspond to paths with large numbers of bounces in the "direct" version of the propagator. This is expected when $\bar{a} \to 0$ (i.e. $\rho \to 0$ since $\rho \sim \frac{\bar{a}^2}{\delta z}$ and δz never becomes large) since having narrow lanes is likely to result in paths with many bounces.

9.3.7 Computation

It is desirable to evaluate the infinitesimal propagator numerically. If the calculations can be performed for sufficiently large and small values of ρ then the asymptotic predictions may be tested. The "direct" version, $\bar{K}_1$, has a complicated functional form. Rather than considering the whole of the expression for $\bar{K}_1$, only part of it (J_1) is considered as a first approach. The formula for J_1 is obtained by replacing the factor $\left(1 - \frac{|\Delta|}{A}\right)$ by 1. The reason for this choice is that this part is very much easier to calculate numerically, since it may be written in terms of the standard function

$$\mathrm{w}(z) = e^{-z^2}\mathrm{erfc}(-iz) \tag{9.63}$$

$$\mathrm{erfc}(z) = \frac{2}{\sqrt{\pi}}\int_z^\infty e^{-t^2}dt \tag{9.64}$$

for which there is a NAG library routine (for complex z). Specifically,

$$J_1 = \frac{1}{\bar{a}}\left[\sqrt{\frac{\rho}{\pi i}}e^{-|\Delta|+i\rho|\Delta|^2} + \sum_{n=1}^{\infty}\left(g(1,n) + g(-1,n) - 2\int_{\phi=0}^{1} g(1-2\phi,(n-1)+\phi)d\phi\right)\right] \tag{9.65}$$

where

$$g(C,N) = \frac{1}{4N}e^{-|\Delta|+i\rho(C+2N)^2|\Delta|^2}\mathrm{w}\left(\sqrt{i}\left[(C+2N)\sqrt{\rho}N + \frac{i}{4\sqrt{\rho}N}\right]\right) \tag{9.66}$$

This should mean that each term in the sum is "quick" to calculate. Also the convergence of the sum in J_1 is not expected to be significantly different, for numerical purposes, from that for the full expression. It is the ratio of the term to the partial sum that is important for truncation purposes. So calculating just the simplest part of $\bar{K}_1$ should be much quicker than considering the full expression. If this is computationally intensive then it is likely that calculation of $\bar{K}_1$ is not worthwhile.

Also, the formulae are the same when $\Delta = 0$ so the results should be numerically correct for this special case. The form of J_1 for large ρ is $e^{-|\Delta|+i\rho|\Delta|^2}$ which is different from that

for $\bar{K}_1$ but the feature that the "zero bounce" term dominates is the same. This is shown by storing the number of terms before truncation in the computation of the sum for a specified precision. Some sample results for the modulus of $\bar{K}_1$ are shown in figure 9.4. In fact, when ρ is not large the number of terms required increases considerably. For $\rho < 10^{-2}$ this method is not practical. The "transformed" version J_p (i.e. equation (9.34) with the factor $\left(1 - \frac{|\Delta|}{A}\right)$ replaced by 1) is more suitable for small values of ρ but this no longer has the advantage of being easy to calculate compared to the corresponding full expression $\bar{K}_p$. However, it is desirable to check the values calculated using the expression J_1. This is indeed possible over a range of values of ρ from 10^{-2} to 10. The problem with the calculation of J_p is the nature of the integration required. The integrand oscillates (at non-constant frequency) and its envelope changes in magnitude within the interval of integration. The way that this occurs makes it difficult to obtain reliable results when ρ is small ($< 10^{-2}$) for a range of values of Δ. However, it is possible to confirm that the $|\Delta|$ dependence does become less apparent as ρ decreases.

The case $\Delta = 0$ is simpler. It is possible to compute values of $\bar{K}_p(\Delta = 0)$ using a modified form of equation (9.42), for values of ρ as small as 10^{-7}. It is necessary to scale the values by a large factor during the calculation in order to avoid computational problems with small numbers. The logarithm of the result is calculated and the factor removed by subtraction. These results can then be compared with the asymptotic values obtained using equation (9.61), i.e. figure 9.5. The graph shows good agreement between the numerical and asymptotic results for small values of ρ. The agreement improves as ρ decreases (as would be expected for an asymptotic formula) until $\log \rho = -15$. For smaller values of ρ the numerical result becomes unreliable.

Apart from the comparisons with asymptotic results, the computational investigation shows that the averaging necessary to produce a well behaved function makes the 1D propagator challenging to calculate. It seems that numerical composition of the 2D propagator is not a practical proposition.

9.3.8 Summary

Reviewing progress in this part of the chapter (i.e. the introduction of randomness into the mirror planes model), it is noted that attention has been restricted to a one-dimensional single stage propagator. It is straightforward to extend this to a 2D single stage propagator since the motion in the direction parallel to the mirror plane strips is unconstrained and so may be included using a factor based on the free space propagator.

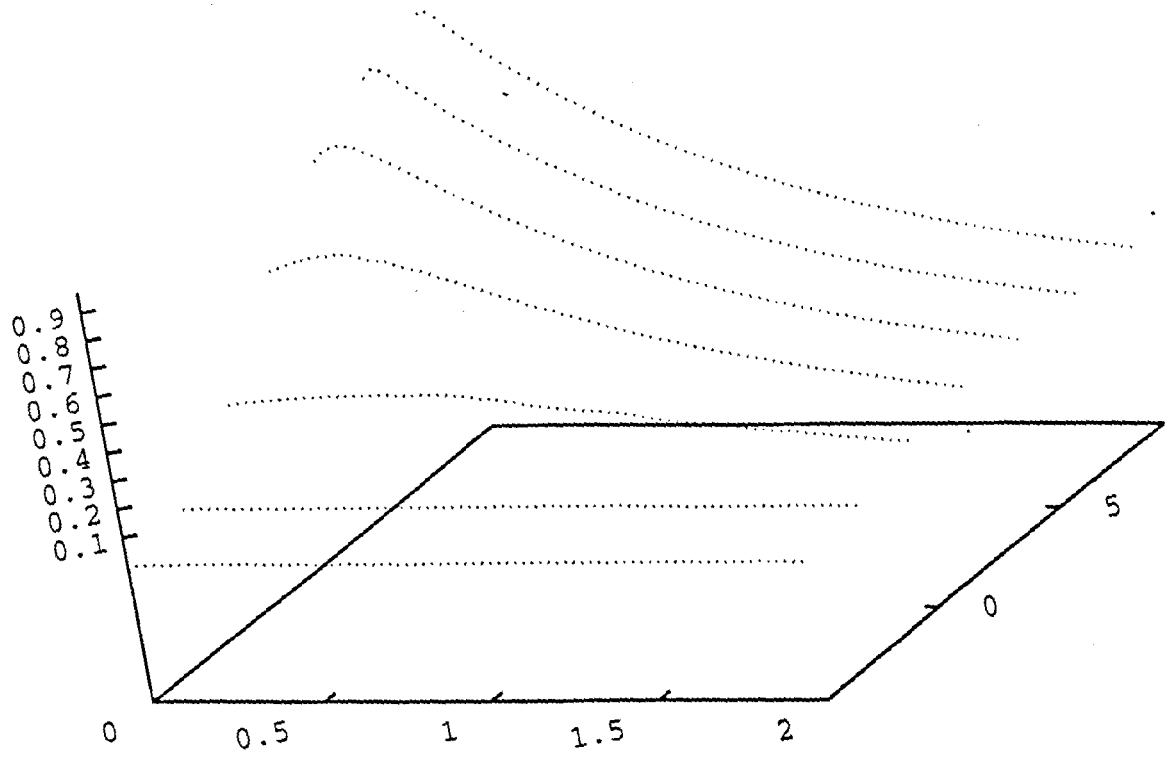


Figure 7.4: $|\tilde{K}_1|$ against Δ and $\log\rho$, values of ρ are from 0.01 to 10^4

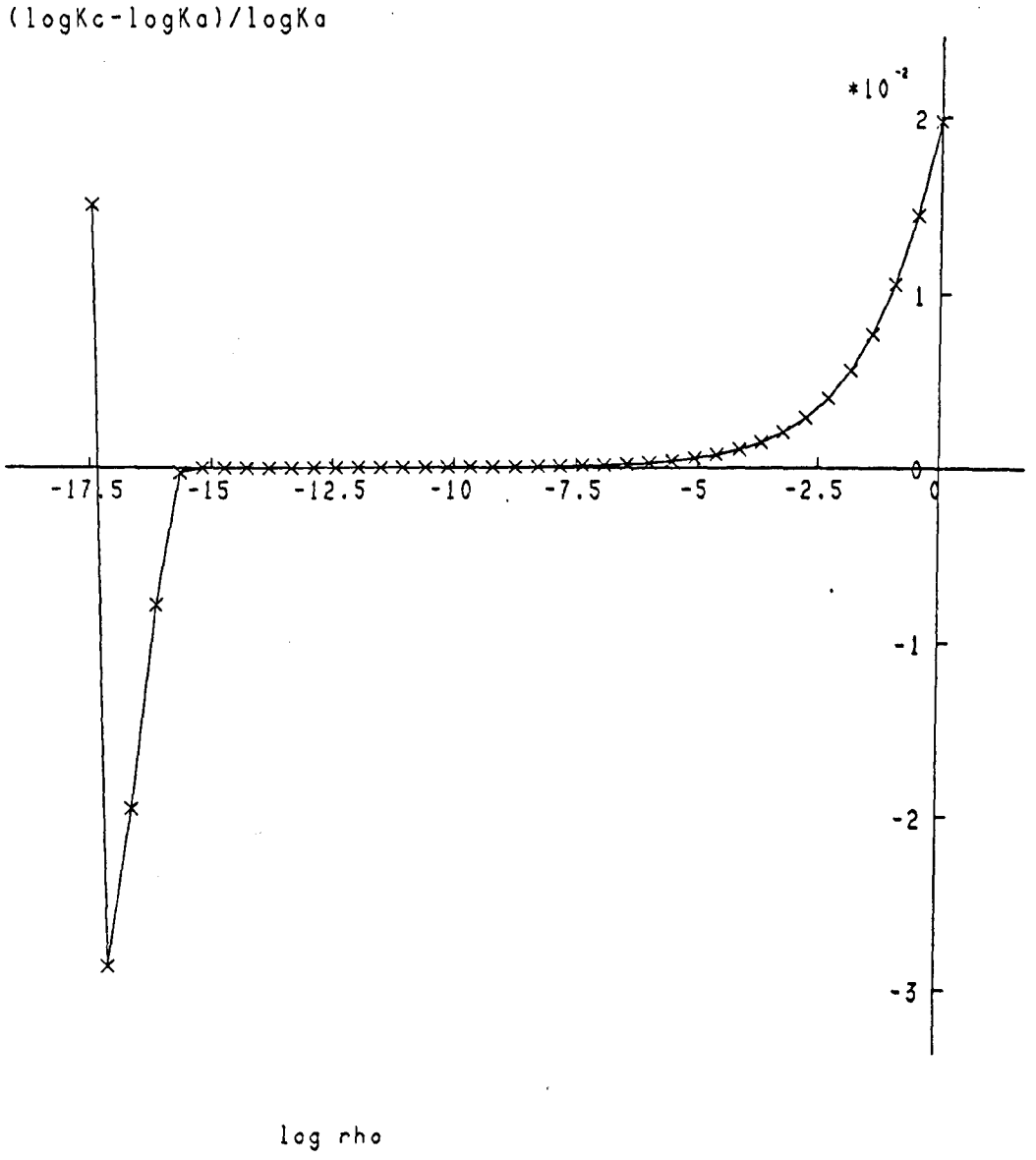


Figure 7.5: Comparison of asymptotic (K_a) and computed (K_c) results for $|\bar{K}_p(\Delta = 0)|$ as a function of ρ

Apart from removing sensitive dependence on initial conditions, averaging over a random distribution of lane widths improves the convergence of the sums. The magnitude of the contribution from paths with large gradients is decreased. This is desirable since large "gradients" correspond to the component of the velocity in the $\underline{n}$ direction (i.e. the "constraint direction") being large (i.e. the constraint being violated to a large extent). Unfortunately, the formulae become considerably more complicated when averaging is included.

The composition of stages to give a finite propagator is not achieved. The analytical formulae are too complicated for this to be feasible for the general case. Consequently, limiting cases are considered (section 9.3.5) in the hope that these will be simple enough to perform composition at least for some special cases. Unfortunately, the resulting expressions are still complicated except when the displacement for the (infinitesimal) stage (i.e. Δx) is taken to be zero. Whilst the $\Delta x = 0$ case does provide an idea of the behaviour to expect, composition would require an expression with the full Δx dependence.

As far as the numerical investigation is concerned, it seems that it is not really practical to calculate the full propagator (i.e. even for a single stage). Although, as a check of (self) consistency, the numerical results do lend support to the asymptotic calculations.

9.4 Summary

This chapter provides an investigation of the effect of "randomness" on models introduced previously. The first part of the chapter (section 9.2) follows on from chapter 8 by modifying the phase screens model. The second part of the chapter (section 9.3) modifies the mirror planes model (of chapter 6).

The investigations in this chapter suggest that using a random, rather than a regular, distribution of "lane widths" (a "lane" is the space between an adjacent pair of mirror planes or the corresponding region in the phase screens model) is a sensible approach in principle — although it can make the formulae more complicated in some cases.

A definite connection can be made between the two parts of the chapter. There is a limit in which the "random phase screens" and the "random mirror planes" results for the averaged single stage propagator coincide (section 9.3.6). Specifically, it is found that as $\Delta \to 0$ the mirror planes (single stage) result tends to the corresponding phase screens result (which is calculated for general ρ). For the types of path segment likely to be of interest, taking the limit $\rho \to \infty$ (expected to be the nonholonomic limit) is sufficient to

ensure that $\Delta \rightarrow 0$ (section 9.3.6). So for the "random" versions of "mirror planes" and "phase screens" there is explicit agreement under the conditions given in chapter 8 as the requirements for the original (i.e. "non random") version of "phase screens" to enforce the constraints to a good approximation. This demonstrates that (as expected) the same criteria are relevant for the random version. The fact that "phase screens" does not (in the general case) enforce the constraints exactly can be illustrated by considering a "test" section of Feynman path (i.e. with a phase associated with it) : if such a path section crosses a lane boundary it simply "picks up" an extra phase of π. For "phase screens", the single stage propagator from a given initial point to a final point in a different lane is not zero as it would be in an (exact) implementation of the mirror planes model.

In previous chapters (section 6.2 for example) it was anticipated that the parameter $\frac{\bar{a}}{\delta z}$ (or $\frac{\bar{a}}{\epsilon}$) would be important. This was based solely on consideration of the structures introduced to enforce the constraint. In this chapter, the parameter ρ has arisen naturally: it contains quantities associated not just with the constraints but also with the "kinetic" part of the problem. If the other factors in ρ are considered constant, then $\rho \sim \frac{\bar{a}^2}{\delta z}$. If the "regime" is to be characterised according to whether the parameter tends to infinity or tends to zero, then there are some cases which will be classified differently depending upon whether $\frac{\bar{a}}{\delta z}$ or $\frac{\bar{a}^2}{\delta z}$ is used as the parameter. Specifically, if $\bar{a} \sim \delta z^p$ as $\delta z \rightarrow 0$ (where $p > 0$) then the "extreme cases" i.e. $p < \frac{1}{2}$ and $p > 1$ are unaffected by the choice of parameter but for "intermediate" cases $\frac{1}{2} < p < 1$ there will be a difference. The reason for this is that using ρ as the parameter takes into account the characteristics of the path as well as the constraint structure. In fact ρ is really the product of two parameters i.e. the "geometric" parameter $\frac{\bar{a}}{\delta z}$ (which describes the "constraint structure") and a parameter determining the classical (i.e. "ray like" in optics) or quantum (i.e. "wave like") nature of the path segments. Provided the "classical limit" is not being taken, for example, then it is appropriate to use $\frac{\bar{a}^2}{\delta z}$ as an "overall" parameter (i.e. for the "interaction" of the path segments with the constraint structure)

Choosing optics to investigate the nature of ρ, it is expected that diffraction (i.e. wave) effects will be important when the wavelength is of the same order as (or larger than) the typical length scale. Taking $\bar{a}$ to be the "typical length scale" suggests using $\frac{\bar{a}}{\lambda}$ as a parameter to distinguish between the wave and ray limits (i.e. $\frac{\bar{a}}{\lambda} \rightarrow \infty$ for the ray limit). Inspection of the formula

$$\rho = \frac{k\bar{a}^2}{2\delta z} \tag{9.67}$$

reveals that ρ does indeed appear to be the product of a "wave/ray" parameter ($\frac{\pi a}{\lambda}$) and a "geometric" parameter ($\frac{\bar{a}}{\delta z}$) i.e.

$$\rho = \left(\frac{k\bar{a}}{2}\right)\left(\frac{\bar{a}}{\delta z}\right) \tag{9.68}$$

where $k = \frac{2\pi}{\lambda}$

If ρ is indeed a parameter distinguishing between the "ord " regime and the "vak " regime, then taking the ray (classical) limit ($k\bar{a} \rightarrow \infty$) seems to "favour" the ordinary nonholonomic (i.e. $\rho \rightarrow \infty$) case (and indeed no vakonomic classical mechanical systems have been observed experimentally). Similarly, the wave (quantum) limit seems to "favour" the vakonomic case (it is relatively straightforward to implement a "vakonomic quantum mechanics" in appendix C). From considering examples of nonholonomic constraints in classical mechanics it seems likely that there will indeed be a "constraint length scale".

Perhaps the simple model system studied here has the same qualitative behaviour as real mechanical systems. In order to find out if this is true, it is necessary to obtain some sort of propagator for a finite time interval in a tractable form. The next chapter presents a new approach to this problem.

Chapter 10

Nonholonomic propagation

10.1 Introduction

The approach to the investigation of propagation over a finite time interval (i.e. "finite propagation") used in previous chapters has generally been to separate the problem into two parts: the consideration of infinitesimal stages (i.e. the time interval of the stage is infinitesimally small); the combination of infinitesimal stages to give a finite time interval. Often attention has been confined to the first part of the process, i.e. exploration of an infinitesimal stage. In this chapter, the aim is to take a more direct approach to the problem of obtaining the "finite propagation" versions of quantities of interest (which avoids the difficulties associated with explicitly combining an infinite number of infinitesimal stages). It will still be beneficial to imagine the continuous case as a limit of many infinitesimal stages, but the transition to a finite interval of the continuous case will be made at the conceptual level and formulae will be written directly for finite propagation. As explained in the following section (i.e. section 10.2) it is the "nonholonomic limit" which is most amenable to this treatment.

10.2 Preliminaries

The idea of the mirror planes model as a method of enforcing the exact constraints (rather than an approximate version) was introduced in chapter 6. This provides the starting point from which the model to be used in this chapter is developed.

As the "nonholonomic limit" $\frac{\bar{a}^2}{\delta z} \to \infty$ (the "nonholonomic limit" is discussed in section 9.4) is approached, the length of a single stage in the z-direction (i.e. δz) becomes small compared to the (average) "lane width" (i.e. $\bar{a}$), it is easy to imagine the mirror

planes (of chapter 6 and section 9.3) becoming like "wires". Conversely, as the "vakonomic limit" is approached, δz becomes large compared to $\bar{a}$ (although both δz and $\bar{a}$ tend to zero) and it is more tempting to think of "strips" in this case. If attention is restricted to the "nonholonomic limit" (which is of particular interest) and the "image planes" (i.e. reflections of the mirror planes in each other) are included, each stage produces a *set of wires.* These are like diffraction gratings. A way of modelling the qualitative effect of a diffraction grating (i.e. spreading light perpendicular to the "grating lines") is to have a "phase screen" with a refractive index which is very rapidly varying in the direction perpendicular to the grating rules (the $\underline{n}$ direction) and not varying in the direction parallel to them (the $\underline{u}$ direction). Previous difficulties caused by periodicity suggest that the refractive index should be (rapidly) randomly varying. A varying refractive index produces an "acceleration" (curvature of the path), so this configuration gives acceleration in the $\underline{n}$ direction (where $\underline{n}$ is the constraint normal vector) but none in the $\underline{u}$ direction (i.e. perpendicular to $\underline{n}$), just as in classical (ordinary) nonholonomic mechanics.

This use of phase screens (to be implemented in section 10.3) differs from the way that they were applied in previous chapters. Previously there were a pair of phase screens associated with each stage: one at the beginning and one at the end of the stage (this approach will also be used in section 10.4). In the new approach there is only a single phase screen associated with each stage. Also, the refractive index is now a continuous, smooth function of transverse (i.e. perpendicular to the z-direction) position (specifically in the $\underline{n}$ direction) whereas previously there were "jumps": the phase change was either 0 or π.

10.3 Random refractive index

Qualitatively, the idea of using a random refractive index seems promising. In order to begin a more *quantitative* investigation, it is necessary to make some choices about the properties of the refractive index (n). In fact, it is convenient to work in terms of the refractivity, φ, which is related to the refractive index by

$$\varphi = \mathrm{n} - 1 \tag{10.1}$$

Taking φ to be a zero-mean Gaussian random function [13] of (transverse) position is a suitable choice since its statistics are then fully defined once its covariance function is specified: it would be difficult to justify the individual specification of higher moments at

this level of modelling. There is no reason to think that φ should depend upon absolute, rather than relative, ("transverse") positions, so it is expedient to take φ to be statistically stationary. A "Markov model" is commonly adopted since it simplifies the mathematics considerably, it is natural to employ one here: it corresponds to the assumption that any two phase screens ("diffraction gratings") are statistically independent, which introduces a delta function (i.e. $\delta(z_2 - z_1)$) into the covariance function.

With these assumptions, the covariance function may be written

$$\langle \varphi(\underline{r}', z)\varphi(\underline{r}'', z + \Delta z)\rangle = v(z)\varrho(\underline{\xi}, z)\delta(\Delta z) \tag{10.2}$$

where

$\Delta z = z_2 - z_1$

$\underline{\xi} = \underline{r}'' - \underline{r}'$

$$\varrho(\underline{\xi}, z) = \frac{\int_{-\infty}^{\infty}\langle \varphi(\underline{r}, z)\varphi(\underline{r} + \underline{\xi}, z + u)\rangle du}{v(z)} \tag{10.3}$$

is a dimensionless autocorrelation function

and

$$v(z) = \int_{-\infty}^{\infty}\langle \varphi(\underline{r}, z)\varphi(\underline{r}, z + u)\rangle du \tag{10.4}$$

is a factor which ensures dimensional consistency

Equation (10.2) is to be interpreted as the "continuous analogue" of

$$\langle \varphi_i(\underline{r}')\varphi_j(\underline{r}'')\rangle = v_i\varrho_i(\underline{\xi})\delta_{ij} \tag{10.5}$$

where δ_{ij} , the Kronecker delta, is 1 when $i = j$ and zero otherwise
and the indices i and j label the phase screens within the "composition". (The "composition" for a finite time interval has an infinite number of phase screens with an infinitely small separation between adjacent phase screens.)

In fact, in the current model, the autocorrelation function ϱ only depends upon one component of $\underline{\xi}$, specifically $\underline{n}.\underline{\xi}$ (where $\underline{n}$ is the "constraint " normal vector)
e.g.

$$\rho(\underline{\xi}, z) = e^{-\left(\frac{\underline{n}(z).\underline{\xi}}{a}\right)^2} \tag{10.6}$$

It is not instructive to consider $\langle K\rangle$. This is a consequence of the way in which the phase depends upon the random medium. For a zero mean Gaussian random medium, the effect of the medium (on K) after averaging is trivial. It is thus preferable to consider KK^* (the "propagator for intensity" in the optical analogy) since this lacks the phase dependence present in K. The average of KK^* over realisations of the random medium

(i.e. the random medium is the continuum limit of sets of phase screens) is denoted by $\langle KK^* \rangle$.

So,

$$\begin{aligned}
&\langle K(\underline{r}''_L, L; \underline{r}''_0, 0) K^*(\underline{r}'_L, L; \underline{r}'_0, 0) \rangle \\
&= \int_{\underline{r}''_0,0}^{\underline{r}''_L,L} \int_{\underline{r}'_0,0}^{\underline{r}'_L,L} \langle e^{ik(S[\underline{r}''(z)] - S[\underline{r}'(z)])} \rangle d^\infty \underline{r}'(z)\, d^\infty \underline{r}''(z) \\
&= \int_{\underline{r}_0,0}^{\underline{r}_L,L} \int_{\underline{\xi}_0,0}^{\underline{\xi}_L,L} \langle e^{ik \int_0^L [\dot{\underline{r}} \cdot \dot{\underline{\xi}} + \varphi(\underline{r} + \frac{1}{2}\underline{\xi}, z) - \varphi(\underline{r} - \frac{1}{2}\underline{\xi}, z)] dz} \rangle d^\infty \underline{\xi}(z)\, d^\infty \underline{r}(z)
\end{aligned} \tag{10.7}$$

where

$$\begin{aligned}
\underline{r}(z) &= \frac{1}{2}(\underline{r}''(z) + \underline{r}'(z)) \\
\underline{\xi}(z) &= \underline{r}''(z) - \underline{r}'(z)
\end{aligned} \tag{10.8}$$

Typical values of φ are taken to be small compared to 1

and use has been made of the result $\frac{1}{2}(\dot{\underline{r}}'')^2 - \frac{1}{2}(\dot{\underline{r}}')^2 = \dot{\underline{r}} \cdot \dot{\underline{\xi}}$

It is desired that the functional in the $\underline{r}(z)$ path integral is highly peaked on the classical ord nonholonomic path (i.e. the path satisfying $\ddot{\underline{r}} = -(\dot{\underline{n}} \cdot \dot{\underline{r}})\underline{n}$ where $\underline{n}$ is the constraint normal vector). It is particularly important that this is true in the ray limit "$k \to \infty$" (corresponding to the classical limit "$\hbar \to 0$" of mechanics). In which limit the peak should become very "sharp" (like a δ functional on the classical ord nonholonomic path). Considering the possibility of performing the $\underline{\xi}(z)$ path integral in equation (10.7) shows that this requirement is not satisfied: a suitable velocity (i.e. $\dot{\underline{r}}$) dependent term is absent. In particular, it is possible to evaluate the path integrals in equation (10.7) using the assumption of small $\underline{\xi}$: which is likely to be valid when k can be considered to be "large" (i.e. the wavelength is very much smaller than characteristic length scales in the problem).

i.e. (following [15])

$$\begin{aligned}
&\langle K(\underline{r}''_L, L; \underline{r}''_0, 0) K^*(\underline{r}'_L, L; \underline{r}'_0, 0) \rangle \\
&= \int_{\underline{\xi}_0,0}^{\underline{\xi}_L,L} \int_{\underline{r}_0,0}^{\underline{r}_L,L} e^{ik \int_0^L \dot{\underline{r}} \cdot \dot{\underline{\xi}} dz} \langle e^{ik \int_0^L \underline{\xi} \cdot \underline{\nabla} \varphi(\underline{r}(z), z) dz} \rangle d^\infty \underline{r}(z)\, d^\infty \underline{\xi}(z)
\end{aligned} \tag{10.9}$$

(where $\underline{\nabla}$ is $\frac{\partial}{\partial \underline{r}}$ a 2D vector operator in the plane perpendicular to the z-direction)

which follows from the last expression in equation (10.7) by taking factors independent of φ outside the average and also using the assumption of "small $\underline{\xi}$".

The next step requires the standard result

$$\langle e^F \rangle = e^{\langle F \rangle} e^{\frac{1}{2}\langle (F-\langle F \rangle)^2 \rangle} \tag{10.10}$$

for F proportional to a Gaussian probability distribution. In fact, only the special case of the result for a distribution with zero mean is required at this stage. Also, the fact that the left side of equation (10.2) is independent of $\underline{r}$ is used.
The result is

$$\begin{aligned}&\langle K(\underline{r}''_L, L; \underline{r}''_0, 0) K^*(\underline{r}'_L, L; \underline{r}'_0, 0) \rangle \\ &= \int_{\underline{\xi}_0, 0}^{\underline{\xi}_L, L} \int_{\underline{r}_0, 0}^{\underline{r}_L, L} e^{ik \int_0^L \dot{\underline{r}} . \dot{\underline{\xi}} dz} e^{-\frac{k^2}{4} \int_0^L \underline{\xi}^2 v(z) \nabla^2 \varrho(\underline{0}) dz} d^\infty \underline{r}(z)\, d^\infty \underline{\xi}(z)\end{aligned} \tag{10.11}$$

This can be evaluated by integrating by parts in the first exponent so that the path integral over $\underline{r}(z)$ gives a "delta functional" on $\ddot{\underline{\xi}}$. Hence the path integral over $\underline{\xi}(z)$ can be evaluated to give

$$\begin{aligned}&\langle K(\underline{r}''_L, L; \underline{r}''_0, 0) K^*(\underline{r}'_L, L; \underline{r}'_0, 0) \rangle \\ &= \left(\frac{k}{2\pi L}\right)^2 \exp\left[ik(\underline{r}_L - \underline{r}_0).\left(\frac{\underline{\xi}_L - \underline{\xi}_0}{L}\right) - \frac{1}{4}k^2 \int_0^L \left[\underline{\xi}_0\left(1 - \frac{z}{L}\right) + \underline{\xi}_L \frac{z}{L}\right]^2 v(z) \nabla^2 \varrho(\underline{0}, z) dz\right]\end{aligned} \tag{10.12}$$

10.4 Random vector potential

It seems that the limitations of the model described in the previous section (10.3) might be removed if velocity dependence could be introduced into the action. Using a (random) *vector* potential rather than a scalar potential will introduce a velocity dependent term into the action. Justification for this comes from a generalisation of the phase screens model (chapter 8). So each infinitesimal stage is again considered to have a pair of phase changing screens associated with it (one at the beginning of the stage and one at the end). The "phase screen pair" to be considered in this section differs from those considered in earlier chapters (chapter 8 and section 9.2). Specifically, the configuration to be considered in this section consists of an arbitrary, smoothly varying, phase changing screen followed by a "complementary" screen (figure 10.1). The second screen is complementary to the first in the sense that, in the $\underline{n}$ direction, zero gradient paths (i.e. non-sloping on a graph of component of displacement in the $\underline{n}$ direction against z) have a phase change of 2π, which may be referred to as "zero" phase change.

Previous versions of "phase screens" employed averaging over "shifts" for each infinitesimal stage (with the *relative* position of the members of the phase screen pair "locked"). In the current model a similar effect can be achieved by averaging over realisations of the arbitrary phase screens, provided that the pair is always "complementary". A consequence of averaging is that the magnitude of $\langle KK^* \rangle$ for a given infinitesimal path segment is largest if the phase change associated with the path segment is "zero" (or 2π). The magnitude of $\langle KK^* \rangle$ for "sloping" path segments decreases as the slope of the path segment is increased.

It seems that $\underline{A}.\dot{\underline{r}}$ could provide the basis for a specification of the phase change for this model, if a suitable "vector potential", $\underline{A}$, can be found. It is not difficult to choose $\underline{A}$ such that the requirement for $\underline{A}.\dot{\underline{r}}$ to be zero for "non-sloping" paths is satisfied, for example. However, there are other conditions to be satisfied. In particular, investigation shows that it is not possible to use a vector potential with non-zero mean to give an expression consistent with the classical (ordinary) nonholonomic equations of motion. As an alternative, a vector potential with zero mean and random magnitude in the $\underline{n}$-direction is chosen. The covariance is specified by

$$\langle A_i(\underline{r}', z) A_j(\underline{r}'', z + \Delta z) \rangle = \varpi \delta(\Delta z) n_i n_j f(\underline{\xi}.\underline{n}) \tag{10.13}$$

where

A_i are the components of the vector potential

n_i are the components of the constraint normal vector $\underline{n}$

$i = 1,2$, $j = 1,2$

f is a function decaying from 1 (when its argument is 0) to 0 (at infinity)

and ϖ is the product of a factor which ensures dimensional consistency and a large dimensionless number (it is desired that the fluctuations are large since it is expected that this will be required in order to enforce the constraint).

So

$$\begin{aligned}
&\langle K(\underline{r}''_L, L; \underline{r}''_0, 0) K^*(\underline{r}'_L, L; \underline{r}'_0, 0) \rangle \\
&= \int_{\underline{r}''_0,0}^{\underline{r}''_L,L} \int_{\underline{r}'_0,0}^{\underline{r}'_L,L} e^{ik\int_0^L \frac{1}{2}(\dot{\underline{r}}''^2 - \dot{\underline{r}}'^2)dz} \langle e^{ik\int_0^L (\underline{A}(\underline{r}'').\dot{\underline{r}}'' - \underline{A}(\underline{r}').\dot{\underline{r}}')dz} \rangle d^\infty \underline{r}'(z)\, d^\infty \underline{r}''(z) \\
&= \int_{\underline{r}_0,0}^{\underline{r}_L,L} \int_{\underline{\xi}_0,0}^{\underline{\xi}_L,L} e^{ik\int_0^L \dot{\underline{r}}.\dot{\underline{\xi}}dz} e^{-\frac{k^2}{2}\varpi \int_0^L [2(\dot{\underline{r}}.\underline{n})^2(1 - f(\underline{n}.\underline{\xi})) + \frac{1}{2}(\dot{\underline{\xi}}.\underline{n})^2(1 + f(\underline{n}.\underline{\xi}))]dz} d^\infty \underline{\xi}(z)\, d^\infty \underline{r}(z)
\end{aligned} \tag{10.14}$$

which is obtained by: using result (10.10) in the special case when the mean is zero (i.e.

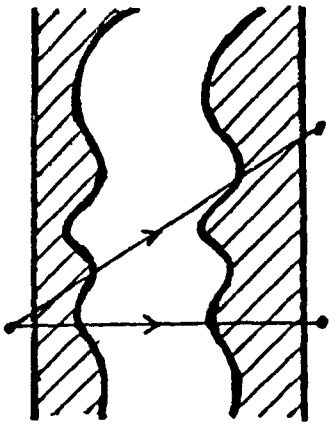


Figure 10.1: Section through a single stage for "generalized phase screens model"

$\langle \underline{A} \rangle = \underline{0}$; writing the squared quantity this introduces into the exponent i.e.

$$\left[\int_0^L (\underline{A}(\underline{r}'').\dot{\underline{r}}'' - \underline{A}(\underline{r}').\dot{\underline{r}}')dz\right]^2$$

as a double integral; expanding the brackets inside the double integral to give four terms; evaluating the "ensemble average" by substituting for the components of the tensors using equation (10.13); reducing the double integral to a single integral using the delta functions introduced in the previous step; transforming to mean and difference coordinates $\underline{r}$ and $\underline{\xi}$ defined in equation (10.8); simplifying the resulting expression by collecting terms together.

It still remains to evaluate the path integrals in equation (10.14). Investigation of this step has not ruled out completely the possibility that a useful result might be obtained. However, it seems to be difficult to obtain a result which satisfies the criteria for success relating to ord nonholonomic mechanical paths becoming important in the classical limit.

10.5 Summary

In this chapter, the scope of the investigation was reduced by excluding consideration of the "vakonomic limit". The objective was to set up a model which allowed the direct calculation of quantities associated with finite propagation. It was found that the requirement that in the classical limit the results should be consistent with the ord nonholonomic (i.e. classical) equations of motion could be used at quite an early stage in the calculations to identify unpromising approaches. Hence the approach based on refractive index was abandoned. It wasn't necessary to perform the calculation described by equations (10.9)-(10.12) although it was included for completeness. An approach based on a vector potential (section 10.4) was attempted instead.

Chapter 11

Conclusions

11.1 The approach

In this thesis an approach has been made to the quantization of mechanical systems subject to nonholonomic constraints. Mostly a special case, with a single nonholonomic constraint in 3D space-time, has been considered. The reason being that this is probably the simplest system which retains the essential features of the problem.

The foremost requirement has been that the classical limit should be correct. This means that the established methods of "constrained dynamics" [33] are not applicable. Although it is artificial to consider classical mechanical systems as "constrained dynamical" systems, it is possible to do so by introducing Lagrange multipliers as extra variables [33]. However, in the case of nonholonomic constraints this leads to a formulation equivalent to vakonomic mechanics. Consequently, quantization based on "constrained dynamics" will have an unphysical (vakonomic) classical limit.

The main focus has been on path integral formulations, although wave equations were also considered to demonstrate that the problems involved are fundamental and not an artefact of the method of quantization. Most of the problems fall into two related groups. The first set of problems are concerned with constraining a quantum mechanical system. Such problems will occur irrespective of the type of constraint. This is exemplified by the calculations for a single infinitesimal stage. For the simple system considered here it is true that until the time dependence (or z dependence in the optical analogy) is included there is no distinction between holonomic and nonholonomic systems. When one dimensional propagators are composed this is also a holonomic case. At least two (space) dimensions are required for nonholonomy. Difficulties associated with the propagator having unphysical

dependence on its parameters can be removed from the model by introducing appropriate averaging. Unfortunately, this tends to make computations very involved, even for the simplest of cases. The way the phase of the infinitesimal propagator depends upon position is expected to be responsible for the averaged propagator tending to zero as the propagation interval becomes finite. Intensity may be considered to avoid this problem but, again, this makes calculations more difficult (i.e. $\langle K^*K \rangle$ compared to $\langle K \rangle$).

The second group of problems are specific to nonholonomic (as opposed to holonomic) constraints. For example, the way that infinitesimal stages are composed to produce "nonholonomic propagation". The use, within the model, of a parameter which gives nonholonomic and vakonomic as limiting cases seems promising. This may reflect a separation between the "constraint scale" and the "quantum scale". However, here again, it has proven difficult to produce tangible results.

11.2 Future directions

In chapter 10 the nonholonomic case is considered specifically (and the time dependence of the constraints is included explicitly). Anisotropy is introduced through the statistical properties of a continuous random medium or magnetic field (specifically its associated vector potential). This would be worth further investigation to find out if the anisotropy introduced in this way will in fact favour the "classical fan" (section 1.5.5) as required.

Although attention has been restricted to a simple special case, it is believed that the approach is general enough to allow the extension of any results obtained using it. Another approach would be to explicitly specialize to a particular system, if, for example, it was believed that a given system might be realized experimentally. Alternatively, a particular type of constraint may be considered, indeed, this approach has been taken for "Quantum rolling" [25].

11.3 Quantum rolling

11.3.1 Introduction

This section provides a summary of the ideas behind " quantum rolling". The objective in the work "Quantum rolling" [25] is to study some examples of discrete systems subject to a map representing a single step of a "rolling" process. The matrix of transition "amplitudes" between states is obtained. The system studied which is of most interest from the current

perspective is a cube "rolling" on an infinite plane (or square grid). Each step of the "rolling" process involves a "flip" of the cube — pivoting about one of its edges in contact with the plane. If the cube is subjected to a series of "flips" which returns it to its starting position, then its orientation will not in general be the same as the initial orientation. This system is non-integrable. A ball rolling without slipping on a plane, as described in section 1.3 shows this effect as well. An example of an integrable system is a tetrahedron "rolling" on a triangular grid (its orientation *can* be expressed as a function of its position).

11.3.2 Possible extensions

It is tempting to try and find a series of polyhedra with high symmetry and increasing numbers of sides, which roll in a non-integrable way. As the number of sides increases they become more like a rolling sphere — at least in the sense that the area of a typical side forms a smaller fraction of the total surface area. As an example, consider an icosahedron (20 triangular sides) rolling on a triangular grid. That this is a non-integrable system may be shown by rolling around any point (i.e. the shortest possible circuit). The number of sides of a polygon can always be increased by slicing off the vertices. If this is done to the icosahedron then a truncated icosahedron is formed with 20 hexagonal and 12 pentagonal faces. This will "roll" on a hexagonal grid and taking a path round any given hexagon on the grid will show that the system is non-integrable. This system is like buckminsterfullerene rolling on a graphite plane — although it is not clear whether such a rolling motion would take place in the real physical system. Things could become rather complicated if a pentagonal face came into contact with the plane during rolling. This does not seem to be a promising way to take a limit.

11.3.3 Physical considerations

Considering instead the classical sphere rolling on a rough plane, the nonholonomic constraint is the no-slip condition. This depends upon microscopic irregularities on the surface of the sphere and the plane "inter-locking" to prevent relative motion. Although these irregularities are small on the "classical scale", they are large on the "quantum scale". It seems that it may not be possible to take this classical nonholonomic constraint to the quantum level without changing its meaning. It is possible that similar problems may in fact occur for all classical nonholonomic systems.

11.4 Models

In the quest to quantize "nonholonomic systems" it has been found that the most direct mathematical approaches to the problem of applying the constraints (e.g. appendix C) do not give expressions with the correct classical limit.

A "modelling approach" is taken to the search for a mechanism for enforcing the constraints. A model system is set up with features which will hopefully enforce the constraints (to a greater or lesser extent). This model is then investigated to find out if its properties are suitable. The model is modified as a result of this investigation and the investigation stage repeated (alternatively it may be rejected as unsuitable). Iteration of this process gives the opportunity to deduce the important features of the way nonholonomic constraints should be enforced.

11.5 Results

Investigation of "the quantization of nonholonomic systems" has led to a number of possible lines of attack by which the problem might be solved being closed. Insight has been gained into the nature of the problem which turns out to be more subtle than might first be imagined.

11.6 Summary

Despite the current interest in the classical mechanics of nonholonomic systems, questions of quantization (for ordinary mechanical nonholonomic systems) have been largely ignored. The path integral based approach to the quantization of nonholonomic systems investigated in this thesis is believed to be unique.

Appendix A

Mechanical principles

A.1 d'Alembert's principle

A mechanical system subject to constraints can be described by Newton's law of motion

$$\frac{d}{dt}(m\dot{\underline{r}}) = \underline{F}_e + \underline{F}_c \tag{A.1}$$

where

$\underline{F}_e$ is the "external" force,

$\underline{F}_c$ is the "force of constraint"

and the corresponding multiple particle equation is obtained by introducing an index i throughout and summing over i.

The "forces of constraint" are generally unknown a priori, as the constraints are stated in terms of constraint equations such as

$$f(\dot{\underline{r}}, \underline{r}, t) = 0 \tag{A.2}$$

Consequently, it is desirable to find a formulation of mechanics which does not require explicit knowledge of the constraint forces. D'Alemberts principle, which is a generalization of the principle of virtual work from statics to dynamics, achieves this goal. A "virtual displacement" is made by changing the configuration of the system by an infinitesimal amount $\delta\underline{r}$ at the instant t. For a system in equilibrium the "virtual work" done in such a displacement is guaranteed to be zero by the principle of virtual work, so (for m = *constant*)

$$(\underline{F}_e + \underline{F}_c - m\ddot{\underline{r}}).\delta\underline{r} = 0 \tag{A.3}$$

If the virtual displacements are restricted to those satisfying the constraints, then the virtual work done by the forces of constraint (i.e. $\underline{F}_c.\delta\underline{r}$) will vanish. For kinematic

constraints

$$\underline{n}(\underline{r},t).\dot{\underline{r}} + n_t(\underline{r},t) = 0 \tag{A.4}$$

the condition on the virtual displacements is

$$\underline{n}(\underline{r},t).\delta\underline{r} + n_t(\underline{r},t)\delta t = 0 \tag{A.5}$$

with $\delta t = 0$ since the virtual displacement is made at a given time t.

So if

$$\underline{n}(\underline{r},t).\delta\underline{r} = 0 \tag{A.6}$$

then equation (A.3) becomes

$$(\underline{F}_e - m\ddot{\underline{r}}).\delta\underline{r} = 0 \tag{A.7}$$

This equation is in the form most commonly used to express d'Alembert's principle. However, since

$$\underline{F}_c.\delta\underline{r} = 0 \tag{A.8}$$

for $\delta\underline{r}$ arbitrary except for the requirement that condition (A.6) is satisfied, the equation of motion can be written as

$$m\ddot{\underline{r}} = \underline{F}_e + \lambda\underline{n} \tag{A.9}$$

where λ is undetermined.

Equation (A.9) can be considered to be a direct consequence of d'Alembert's principle.

This equation and the constraint equations taken together can be used to determine the motion without explicit reference to the forces of constraint

A.2 Gauss's principle of least constraint

At a certain time, t, a system has a prescribed configuration and velocity. The objective is to find equations to determine the acceleration.

For a motion with acceleration $\ddot{\underline{r}}$, differentiating the equation of constraint

$$\underline{n}.\dot{\underline{r}} + n_t = 0 \tag{A.10}$$

provides a condition on $\ddot{\underline{r}}$, i.e.

$$\underline{n}.\ddot{\underline{r}} + \dot{\underline{n}}.\dot{\underline{r}} + \dot{n}_t = 0 \tag{A.11}$$

Considering another possible motion with the same configuration and the same velocity at time t, but with acceleration $\ddot{\underline{r}} + \Delta\ddot{\underline{r}}$, the equation corresponding to (A.11) is

$$\underline{n}.(\ddot{\underline{r}} + \Delta\ddot{\underline{r}}) + \dot{\underline{n}}.\dot{\underline{r}} + \dot{n}_t = 0 \tag{A.12}$$

Equations (A.11) and (A.12) together give

$$\underline{n}.\Delta\ddot{\underline{r}} = 0 \tag{A.13}$$

Taking the components of equation (A.9) ("d'Alembert's principle") in the direction of $\Delta\ddot{\underline{r}}$ and using equation (A.13) gives

$$(\underline{F}_e - m\ddot{\underline{r}}).\Delta\ddot{\underline{r}} = 0 \tag{A.14}$$

Since the configuration and velocity are considered as constants, this may be written

$$(\underline{F}_e - m\ddot{\underline{r}}).\delta(\underline{F}_e - m\ddot{\underline{r}}) = 0 \tag{A.15}$$

In equation (A.15) attention has been restricted from the finite changes in equation (A.14) to the special case of infinitesimal variations.

Equation (A.15) means that

$$\delta Z = 0 \tag{A.16}$$

for

$$Z = \frac{1}{2m}(\underline{F}_e - m\ddot{\underline{r}})^2 \tag{A.17}$$

where Z is considered as a function of $\ddot{\underline{r}}$.

Gauss called the quantity Z the "constraint" of the motion and expressed equation (A.16) as the "principle of least constraint": the actual motion occuring in nature is such that under the given kinematic conditions (i.e. equation (A.10) in this case) the "constraint" becomes as small as possible.

The position and velocity components are constants (they were specified initially) so Z is a quadratic function (with constant coefficients) of the acceleration components.

Hertz's geometrical interpretation ("principle of straightest path") of Gauss's principle of least constraint for the special case of no external forces is given in section 1.5.3.

A.3 Quasi-coordinates

There is an explicit functional relationship between generalized coordinates (q_i) and physical coordinates. It is convenient, especially when dealing with nonholonomic systems, to use a more general type of coordinates, γ_i. Such a quantity γ is defined by integrating the differential

$$d\gamma = \sum_j a_j(\underline{q}, t)dq_j + a_t(\underline{q}, t)dt \tag{A.18}$$

along the trajectory of the system from the point $(\underline{q}_0, t_0)$ to an arbitrary point $(\underline{q}, t)$. If the differential (A.18) is an exact differential, then the quantity γ will be a function of $\underline{q}$ and t only: it will not depend on the path taken to reach this point. In this case γ could be used as a coordinate for the system. If the differential (A.18) is not an exact differential then this will not be possible: the quantity γ will depend upon the path (as well as $\underline{q}$ and t). A quantity, γ, obtained by integrating a differential along a trajectory in this way is called a quasi-coordinate and the quantity $\dot{\gamma}$ is called a quasi-coordinate velocity. Clearly the coordinate system used should be able to specify the configuration of the system. It is for this reason that a sufficient number of the original coordinates are usually retained in practical examples.

A.4 The Gibbs-Appell equations

The Gibbs-Appell equations are closely related to Gauss's principle. They are usually stated in terms of quasi-coordinates. The flexibility provided by quasi-coordinates means that the Gibbs-Appell equations are prefered over the explicit use of Gauss's principle for the solution of all but the simplest of problems (although recently a method for applying Gauss's principle without the explicit use of quasi-coordinates has been expounded [34]).

To show the relationship between the Gibbs-Appell equations and Gauss's principle, it is desirable to express Gauss's "constraint", Z, in terms of quasi-coordinates. The first step is to write Z in index notation (since quasi-coordinates are to be used) i.e.

$$Z = \frac{1}{2}\sum_{j=1}^{N} m_j \left(\ddot{x}_j - \frac{F_j}{m_j}\right)^2 \tag{A.19}$$

where N is the dimension of coordinate space

and $m_j = m \forall j$ is possible for simple systems.

The accelerations $\ddot{x}_j$ in equation (A.19) now need to be considered as a function of the $\ddot{\gamma}_i$. This can be achieved by using the equation

$$\ddot{x}_j = \sum_{i=1}^{k} \alpha_{ji}\ddot{\gamma}_i + \sum_{i=1}^{k} \frac{d\alpha_{ji}}{dt}\dot{\gamma}_i + \frac{d\alpha_j}{dt} \tag{A.20}$$

$j = 1, \ldots, N$

where $k = N - l$ (l is the number of constraints)

Equation (A.20) is obtained from the relation between the velocity systems

$$\dot{x}_j = \sum_{i=1}^{k} \alpha_{ji}\dot{\gamma}_i + \alpha_j \tag{A.21}$$

$j = 1, \ldots, N$

by differentiation with respect to time. In fact, the quasi-coordinates are usually introduced using equations of the form (A.18) or the "velocity version" of this. It is then neccessary to solve for the $\dot{x}_j$ in order to obtain equations (A.21). The reason for this is that some of the quasi-coordinate velocities are usually defined to be constraints. So, for example, a constraint such as

$$n_x \dot{x} + n_y \dot{y} + n_t = 0 \tag{A.22}$$

could be included by defining a quasi-coordinate velocity as

$$\dot{\gamma} = n_x \dot{x} + n_y \dot{y} + n_t \tag{A.23}$$

Applying the result (A.20) to equation (A.19) gives

$$Z' = G - \sum_{j=1}^{N} \left(\sum_{i=1}^{k} \alpha_{ji} \ddot{\gamma}_i \right) F_j \tag{A.24}$$

where $G = \frac{1}{2} \sum_{j=1}^{N} m_j \ddot{x}_j^2$ is considered as a function of $\ddot{\gamma}$.

In fact Z' differs from Z, but only by terms not containing accelerations (which are unimportant).

It is neccessary to change from the physical components of the force (i.e. F_j) appearing in equation (A.24) to the generalized components of the force (Γ_i) corresponding to the quasi-coordinates. These are defined by the equation for the virtual work

$$\delta W = \sum_{i=1}^{k} \Gamma_i \delta \gamma_i \tag{A.25}$$

Considering also the equation for virtual work in terms of the physical components of the force, i.e.

$$\delta W = \sum_{j=1}^{N} F_j \delta x_j \tag{A.26}$$

and substituting for the δx_j using the equation

$$\delta x_j = \sum_{i=1}^{k} \alpha_{ji} \delta \gamma_i \qquad j = 1, \ldots, N \tag{A.27}$$

which is the relation between virtual displacements corresponding to the relation (A.21) between velocities.

The result is

$$\delta W = \sum_{i=1}^{k} \left(\sum_{j=1}^{N} \alpha_{ji} F_j \right) \delta \gamma_i \tag{A.28}$$

Comparing this with the defining equation (A.25) gives

$$\Gamma_i = \sum_{j=1}^{N} \alpha_{ji} F_j \tag{A.29}$$

which allows equation (A.24) to be written

$$Z' = G - \sum_{i=1}^{k} \Gamma_i \ddot{\gamma}_i \tag{A.30}$$

So Gauss's principle is equivalent to the requirement that $G - \sum_{i=1}^{k} \Gamma_i \ddot{\gamma}_i$ is a minimum for the actual motion. The first order conditions for a stationary value are sufficient, i.e.

$$\frac{\partial G}{\partial \ddot{\gamma}_i} = \Gamma_i \qquad i = 1, \ldots, k \tag{A.31}$$

These are the Gibbs-Appell equations. It is clear that terms in G which do not contain a $\ddot{\gamma}_i$ can be omitted, as far as the equations of motion are concerned.

A.5 Example

If a simple example is considered, then it is possible to compare the form of the Gibbs-Appell equations with the standard result. Considering, for example, a single particle ($m_i = m \forall i$) in 3D space, subject to the single constraint

$$n_x(t)\dot{x} + n_y(t)\dot{y} = 0 \tag{A.32}$$

with $n_x^2 + n_y^2 = 1$

and no external forces, suggests defining the quasi-coordinate velocities

$$\dot{\gamma}_1 = n_x \dot{x} + n_y \dot{y} \tag{A.33}$$

$$\dot{\gamma}_2 = n_y \dot{x} - n_x \dot{y} \tag{A.34}$$

$$\dot{\gamma}_3 = \dot{z} \tag{A.35}$$

although these choices for $\dot{\gamma}_2$ and $\dot{\gamma}_3$ are arbitrary.

The next stage is to write

$$G = \frac{1}{2} m (\ddot{x}^2 + \ddot{y}^2 + \ddot{z}^2) \tag{A.36}$$

in terms of $\ddot{\gamma}_1$, $\ddot{\gamma}_2$ and $\ddot{\gamma}_3$. Then the Gibbs-Appell equations give

$$\ddot{\gamma}_2 = 0 \tag{A.37}$$

$$\ddot{\gamma}_3 = 0 \tag{A.38}$$

and we have

$$\ddot{\gamma}_1 = 0 \tag{A.39}$$

since the derivative of the constraint with respect to time is zero (the constraint holds for all values of time).

So the solution is just as expected from conservation of energy and the constraint equation. i.e.

$$\underline{u}.\dot{\underline{r}} = v \tag{A.40}$$

$$\underline{n}.\dot{\underline{r}} = 0 \tag{A.41}$$

$$\dot{z} = k \tag{A.42}$$

where

$$\underline{n} = (n_x, n_y, 0) \tag{A.43}$$

$$\underline{u} = (n_y, -n_x, 0) \tag{A.44}$$

$$\dot{\underline{r}} = (\dot{x}, \dot{y}, \dot{z}) \tag{A.45}$$

(v and k are constants)

so the kinetic energy $= \frac{1}{2}m\dot{\underline{r}}^2 = \frac{1}{2}mv^2$.

A.6 Discussion

As discussed in the main text (section 1.5.1), a principle of stationary action is required for the path integral quantization which is the main topic of this thesis. Consequently, other mechanical principles (such as Gauss's principle and the Gibbs-Appell equations) are of no direct significance. They are included only for completeness and are placed in an appendix to keep unneccessary diversions out of the main text. Further details can be found in [30, 8, 22, 12]).

Discussion of the principles of classical mechanics is often complicated by the fact that the same name may mean different things to different people. A good example of this is "Hamilton's principle" this is sometimes taken to be the same as the principle of stationary action, but sometimes it is used in a "generalized" sense ([18, 29] for example). This has caused confusion in the past. It is for this reason that the use of the term "Hamilton's principle" has been avoided in the main text. The main focus of this work is mechanics with constraints and the most important methods have been included: d'Alembert's principle,

Gauss's principle and the Gibbs-Appell equations in this appendix; Dirac's method in section 1.6 (and appendix B); variational principles (in the conventional sense i.e. using the calculus of variations) are represented by the principle of stationary action in section 1.5.2. The basic result is unsurprising: a given principle gives the correct equations of motion if it can be "derived" from (shown to be equivalent to) the fundamental principle of mechanics (d'Alembert's principle). In the case of the principle of stationary action this is not possible if the constraints are nonholonomic.

Appendix B

Constrained Hamiltonian systems

B.1 Introduction

The purpose of this appendix is to show that applying Dirac's procedure for passing from the Lagrangian to the Hamiltonian description of classical (constrained) dynamics does not yield the correct equations of motion when the constraints are nonholonomic. Evidence is presented to support the suggestion that the "Dirac" equations of motion are consistent with vakonomic "mechanics". In the holonomic case the equations do correctly describe the observed motion, in the nonholonomic case they do not.

B.2 Equations of motion

The system to be considered is a particle in three space dimensions subject to the constraint

$$\dot{\underline{R}}.\underline{N}(\underline{R}) = 0 \tag{B.1}$$

where $\underline{R} = (x, y, z)$ and $\underline{N}^2 = 1$

This is more general than the special case considered in chapters 5-11. The reason for considering this generalization is that it includes non-trivial holonomic cases (the holonomic case of the simple system considered in chapters 5-11 had a constant normal vector $\underline{n}$).

The Lagrangian is

$$L = \frac{m}{2}\dot{x}^i\dot{x}_i - \lambda N_i\dot{x}^i \tag{B.2}$$

where the $\dot{x}^i$ $(i = 1, 2, 3)$ are the components of $\underline{R}$ (and summation over identical indices is implied)

this is a "singular" Lagrangian only if the multiplier λ is considered to be an additional

coordinate. Making this assumption in order to apply Dirac's procedure and obtaining the generalized momenta gives

$$P_i \equiv \frac{\partial L}{\partial \dot{x}^i} = m\dot{x}_i - \lambda N_i \tag{B.3}$$

$$P_\lambda \equiv \frac{\partial L}{\partial \dot{\lambda}} \approx 0 \tag{B.4}$$

since $P_\lambda (\equiv \chi_1)$ is a primary constraint.

Following the standard procedure [27] , the total Hamiltonian is

$$H_T = \frac{1}{2m}(\dot{P}_i + \lambda N_i)^2 + uP_\lambda \tag{B.5}$$

where u is a multiplier.

The time evolution for χ_1 is obtained from

$$\begin{aligned} \frac{d\chi_1}{dt} &= \{\chi_1, H_T\} \\ &= N^i(P_i + \lambda N_i) \end{aligned} \tag{B.6}$$

using the definition of the Poisson bracket of the system

$$\{A, B\} \equiv \frac{\partial A}{\partial x^i}\frac{\partial B}{\partial P_i} - \frac{\partial B}{\partial x^i}\frac{\partial A}{\partial P_i} + \frac{\partial A}{\partial \lambda}\frac{\partial B}{\partial P_\lambda} - \frac{\partial B}{\partial \lambda}\frac{\partial A}{\partial P_\lambda} \tag{B.7}$$

The consistency condition $\dot{\chi}_1 \approx 0$ gives rise to the secondary constraint

$$\begin{aligned} \chi_2 &\equiv N^i(P_i + \lambda N_i) \\ &= N^i P_i + \lambda \end{aligned} \tag{B.8}$$

This process terminates upon obtaining

$$\dot{\chi}_2 \approx 0 \tag{B.9}$$

The constraints χ_1 and χ_2 are second-class since

$$\{\chi_1, \chi_2\} = -1 \tag{B.10}$$

but it is possible to use the weak equations

$$\chi_1 \approx 0 \tag{B.11}$$

$$\chi_2 \approx 0 \tag{B.12}$$

as strong equations provided that the Poisson bracket is replaced by the Dirac bracket

$$[A, B] \equiv \{A, B\} - \sum_{r,s=1}^{2} \{A, \chi_r\} c_{rs} \{\chi_s, B\} \tag{B.13}$$

where c_{rs} is the inverse of the matrix

$$\begin{pmatrix} \{\chi_1,\chi_1\} & \{\chi_1,\chi_2\} \\ \{\chi_2,\chi_1\} & \{\chi_2,\chi_2\} \end{pmatrix} \tag{B.14}$$

then the Hamiltonian equation (B.5) may be written as

$$H = \frac{1}{2m} P_i(\delta^{ij} - N^i N^j) P_j \tag{B.15}$$

leading to Hamiltonian equations of motion

$$\dot{x}^i = (\delta^{ij} - N^i N^j) P_j \tag{B.16}$$

$$\begin{aligned} \dot{P}_i &= (N^j P_j)\frac{\partial}{\partial x^i}(N^k P_k) \\ &= (N^j P_j) P_k \frac{\partial N^k}{\partial x^i} \end{aligned} \tag{B.17}$$

In the holonomic case the constraint may be integrated i.e. $N_i \dot{x}^i = \dot{f}(x^1, x^2, x^3)$.
So considering a particular surface $f(x^1, x^2, x^3)$ (specified by the initial conditions), the identity

$$\frac{\partial^2 f}{\partial x^i \partial x^j} = \frac{\partial^2 f}{\partial x^j \partial x^i} \tag{B.18}$$

may be written

$$\frac{\partial N_j}{\partial x^i} = \frac{\partial N_i}{\partial x^j} \tag{B.19}$$

Substituting this relation into equation (B.17) gives

$$\dot{P}_i = (N^j P_j) P_k \frac{\partial N^i}{\partial x^k} \tag{B.20}$$

So this equation and equation (B.16) are the equations of motion in the holonomic case, they reduce [27] to

$$\ddot{x}^i + n^i n_{j,k} \dot{x}^j \dot{x}^k = 0 \tag{B.21}$$

which is the standard equation of motion for a holonomic system [27]. So the Dirac procedure gives the correct classical equations of motion when the constraints are holonomic. The question is, what do the equations of motion that the Dirac procedure gives for non-holonomic constraints represent, i.e.

$$\dot{\underline{R}} = (1 - \underline{N}\,\underline{N})\underline{P}(\underline{R}) = 0 \tag{B.22}$$

$$\begin{aligned} \dot{\underline{P}} &= (\underline{N}.\underline{P})\underline{\nabla}(\underline{N}.\underline{P}) \\ &= (\underline{N}.\underline{P})(\underline{P}.\underline{\nabla})\underline{N} + (\underline{N}.\underline{P})(\underline{P} \times (\underline{\nabla} \times \underline{N})) \end{aligned} \tag{B.23}$$

Lagrange multipliers are used as extra coodinates in the Dirac procedure. The same is true in vakonomic "mechanics". This suggests that the equations (B.22) and (B.23) may be related to the vakonomic equations of motion. The method in [2] for passing from L to H in vakonomic mechanics may be used to investigate this conjecture. Thus, to put equations (1.11) and (B.1) in Hamiltonian form, introduce the canonical momenta

$$\begin{aligned} \underline{P} &= \frac{\partial L}{\partial \dot{\underline{R}}} + \lambda \frac{\partial}{\partial \dot{\underline{R}}}(\underline{N}.\dot{\underline{R}}) \\ &= m\dot{\underline{R}} + \lambda \underline{N} \end{aligned} \tag{B.24}$$

Considering both $\dot{\underline{R}}$ and λ to be "solved for" in terms of $\underline{P}$ and $\underline{R}$, the Hamiltonian is obtained using

$$\begin{aligned} H &= \dot{\underline{R}}.\underline{P} - L \\ &= \frac{1}{2m}(\underline{P} - \lambda \underline{N})^2 + \lambda \underline{N}.(\underline{P} - \lambda \underline{N}) \\ &= \frac{\underline{P}^2}{2m} - \frac{\lambda^2}{2m} \end{aligned} \tag{B.25}$$

Equation (B.24) and the constraint, equation (B.1), give

$$\lambda = \underline{P}.\underline{N} \tag{B.26}$$

So the equations of motion are

$$\begin{aligned} \dot{\underline{R}} &= \frac{\partial H}{\partial \underline{P}} \\ &= \frac{1}{m}\left(\underline{P} - \lambda \frac{\partial \lambda}{\partial \underline{P}}\right) \\ &= \frac{1}{m}(\underline{P} - (\underline{N}.\underline{P})\underline{N}) \end{aligned} \tag{B.27}$$

$$\begin{aligned} \dot{\underline{P}} &= -\frac{\partial H}{\partial \underline{R}} \\ &= \frac{\lambda}{m}\frac{\partial \lambda}{\partial \underline{R}} \\ &= \frac{1}{m}(\underline{P}.\underline{N})\nabla(\underline{P}.\underline{N}) \end{aligned} \tag{B.28}$$

which are the same as equations (B.16) and (B.17). This agreement between the equations of motion resulting from the Dirac method and those from the Hamiltonian form of vakonomic mechanics supports the conjectured link between the two approaches

To make the connection more explicit the vakonomic equations will be solved in the original Lagrangian form and the solution shown to be the same as that obtained by integrating Dirac's Hamiltonian equations of motion. This procedure will be carried out

for a simple special case with $\underline{N} = \underline{N}(z)$ only and $N_z = 0$,i.e. $\underline{N} = (\underline{n}(z), 0)$. Similarily, $\underline{R}$ is written $\underline{R} = (\underline{r}, z)$. In this notation the Dirac equations of motion (B.16) and (B.17) become

$$\dot{\underline{r}} = \frac{1}{m}(1 - \underline{nn})\underline{p} \tag{B.29}$$

$$\dot{z} = \frac{1}{m}p_z \tag{B.30}$$

$$\begin{aligned} \dot{\underline{p}} &= \frac{1}{m}(\underline{n}.\underline{p})\nabla(\underline{n}.\underline{p}) \\ &= 0 \end{aligned} \tag{B.31}$$

$$\dot{p}_z = \frac{1}{m}(\underline{n}.\underline{p})\frac{\partial}{\partial z}(\underline{n}.\underline{p}) \tag{B.32}$$

Integrating equation (B.31) gives

$$\underline{p} = \underline{c}_d \tag{B.33}$$

where $\underline{c}_d$ is a constant vector.

Substituting (B.33) and (B.30) into (B.32) gives

$$\begin{aligned} \ddot{z} &= \dot{z}\frac{d\dot{z}}{dz} \\ &= \frac{1}{m^2}(\underline{n}.\underline{c}_d)\left(\frac{d\underline{n}}{dz}.\underline{c}_d\right) \end{aligned} \tag{B.34}$$

or

$$\frac{1}{2}\frac{d\dot{z}^2}{dz} = \frac{1}{m^2}\frac{1}{2}\frac{d}{dz}(\underline{n}.\underline{c}_d)^2 \tag{B.35}$$

and substituting (B.33) into (B.29) gives

$$\dot{\underline{r}} = \frac{1}{m}(1 - \underline{nn})\underline{c}_d \tag{B.36}$$

Integrating equations (B.35) and (B.36) provides the solution of the equations of motion.

For comparison, the Lagrangian form of the vakonomic equations

$$\frac{d}{dt}(m\dot{\underline{R}} + \lambda\underline{N}) - \lambda\nabla(\underline{N}.\dot{\underline{R}}) = 0 \tag{B.37}$$

$$\underline{N}.\dot{\underline{R}} = 0 \tag{B.38}$$

for this special case may be written

$$\frac{d}{dt}(m\dot{\underline{r}} + \lambda\underline{n}) = 0 \tag{B.39}$$

$$m\ddot{z} - \lambda\left(\frac{d\underline{n}}{dz}.\dot{\underline{r}}\right) = 0 \tag{B.40}$$

$$\underline{n}.\dot{\underline{r}} = 0 \tag{B.41}$$

Integrating (B.39) gives

$$m\dot{\underline{r}} + \lambda \underline{n} = \underline{c}_v \tag{B.42}$$

where $\underline{c}_v$ is a constant vector.

Taking the $\underline{n}$ component of equation (B.42) and using the constraint condition (B.41) gives

$$\lambda = \underline{n}.\underline{c}_v \tag{B.43}$$

substituting this into equation (B.42) gives

$$\begin{aligned} \dot{\underline{r}} &= \frac{1}{m}(\underline{c}_v - (\underline{n}.\underline{c}_v)\underline{n}) \\ &= \frac{1}{m}(1 - \underline{n}\,\underline{n})\underline{c}_v \end{aligned} \tag{B.44}$$

Taking the $\frac{d\underline{n}}{dz}$ component of equation (B.42) gives

$$m\frac{d\underline{n}}{dz}.\dot{\underline{r}} = \underline{c}_v.\frac{d\underline{n}}{dz} \tag{B.45}$$

where use has been made of

$$\underline{n}.\frac{d\underline{n}}{dz} = 0 \tag{B.46}$$

(since $\underline{n}^2 = 1$)

substituting equations (B.43) and (B.45) into equation (B.40) gives

$$m\dot{z}\frac{d\dot{z}}{dz} = \frac{1}{m}(\underline{n}.\underline{c}_v)\left(\frac{d\underline{n}}{dz}.\underline{c}_v\right) \tag{B.47}$$

or

$$\frac{1}{2}\frac{d}{dz}(\dot{z}^2) = \frac{1}{m^2}\frac{1}{2}\frac{d}{dz}(\underline{n}.\underline{c}_v)^2 \tag{B.48}$$

Integrating equations (B.44) and (B.48) provides the solution of the vakonomic equations of motion.

Comparing equations (B.44) and (B.48) with equations (B.35) and (B.36) shows that they are identical if $\underline{c}_v = \underline{c}_d$. This will indeed be the case since

$$\underline{p} = m\dot{\underline{r}} + \lambda \underline{n} \tag{B.49}$$

from the definition of the canonical momentum. Thus there is agreement between the solutions of the equations of motion for Dirac's method and vakonomic "mechanics" for this special case where an explicit solution is attainable.

The simple system considered here is similar to the one considered in the bulk of the thesis except that the z coordinate has not been identified with time, t. This could be achieved, without employing the more complicated explicitly time dependent theory, by including a constraint such as $\dot{z} = 1$.

Appendix C

A first approach to quantization

A natural approach to the problem of quantizing a system subject to nonholonomic constraints using Feynman's path integral formulation, would be to take the unconstrained path integral and then impose the constraint upon it using a "delta functional" i.e.

$$K(\underline{r}_b, t_b; \underline{r}_a, t_a) = \int_{\underline{r}_a, t_a}^{\underline{r}_b, t_b} e^{\frac{i}{\hbar}\int_{t_a}^{t_b} \frac{m}{2}\dot{\underline{r}}^2 dt} \delta[f(\underline{r}, \dot{\underline{r}}, t)] d^\infty \underline{r}(t) \tag{C.1}$$

if the constraint is $f(\underline{r}, \dot{\underline{r}}, t) = 0$ (and $t_b > t_a$).

The "delta functional" $\delta[f]$ is an infinite product of delta functions — one for each "time-slice". If the integral representation

$$\delta(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ixy} dy \tag{C.2}$$

is used for these delta functions, then it is found that the analogous representation for the "delta functional" involves a functional integral, i.e.

$$\delta[f] = \int_{t_a}^{t_b} e^{i\int_{t_a}^{t_b} yf\, dt} d^\infty y(t) \tag{C.3}$$

Using this form in the expression for the propagator, K, gives (for $t_b > t_a$)

$$K(\underline{r}_b, t_b; \underline{r}_a, t_a) = \int_{\underline{r}_a, t_a}^{\underline{r}_b, t_b} \int_{t_a}^{t_b} e^{\frac{i}{\hbar}\int_{t_a}^{t_b} (\frac{m}{2}\dot{\underline{r}}^2 + \lambda f(\underline{r}, \dot{\underline{r}}, t)) dt} d^\infty \lambda(t) d^\infty \underline{r}(t) \tag{C.4}$$

From the standard result for the classical limit (section 2.4.5), the paths which make the action stationary become dominant. In fact, this prescription for obtaining the classical motion is easy to follow: once it is realised that the expression (C.4) for the propagator is exactly what would be obtained if one imposed the constraint on the classical system using Lagrange multipliers and considered these multipliers as additional coordinates. This is exactly what one does in vakonomic "mechanics" before applying the principle of stationary action. In other words the classical limit is vakonomic "mechanics".

Appendix D

Vakonomic solutions for a position independent constraint

D.1 Introduction

This appendix contains analytical results for the 2D (space) vakonomic system in the case when the components of the normal vector to the planelets (in 3D space-time) are independent of position $\underline{r}$, i.e. $\underline{n} = \underline{n}(t)$, $n_t = n_t(t)$ where $\underline{r}$ and $\underline{n}$ are 2D vectors. When $n_t \neq 0$ the planelets are inclined to the t direction.

D.2 Classical

The constraint, $\underline{n}.\dot{\underline{r}} + n_t = 0$, is applied using a multiplier λ

$$\delta \int_0^t \left(\frac{m}{2} \dot{\underline{r}}^2 + \lambda (\underline{n}.\dot{\underline{r}} + n_t) \right) d\tau = 0 \tag{D.1}$$

from variations with respect to $\underline{r}(\tau)$

$$\frac{d}{dt}(m\dot{\underline{r}} + \lambda \underline{n}) = 0 \tag{D.2}$$

whilst variations with respect to $\lambda(\tau)$ give the constraint equation. Integrating (D.2) gives

$$m\dot{\underline{r}} + \lambda \underline{n} = \underline{c} \tag{D.3}$$

where $\underline{c}$ is a constant (vector)

Taking the component of this equation in the $\underline{n}$ direction and substituting for $\underline{n}.\dot{\underline{r}}$ from the constraint equation gives

$$\lambda = \frac{\underline{n}.\underline{c} + m n_t}{n^2} \tag{D.4}$$

where $n = |\underline{n}|$

substituting this expression for $\lambda(\tau)$ into (D.3) and integrating provides $\Delta\underline{r}(t)$ in terms of $\underline{c}$

$$\Delta\underline{r}(t) = \left[\frac{1}{m}\int_0^t \left(1 - \frac{\underline{n}\,\underline{n}}{n^2}\right) d\tau\right] \underline{c} - \int_0^t \frac{n_t}{n^2}\underline{n}\, d\tau \tag{D.5}$$

If the displacement is known at some time, the final displacement $\Delta\underline{r}_T = \underline{r}(T) - \underline{r}(0)$ for example, then this equation determines $\underline{c}$

$$\underline{c} = \left[\frac{1}{m}\int_0^T \left(1 - \frac{\underline{n}\,\underline{n}}{n^2}\right) d\tau\right]^{-1} \left(\Delta\underline{r}_T + \int_0^T \frac{n_t}{n^2}\underline{n}\, d\tau\right) \tag{D.6}$$

so that $\Delta\underline{r}(t)$ is now determined in terms of $\Delta\underline{r}_T$

To obtain the "classical action"

$$S_{cl} = \int_0^t \frac{m}{2}\dot{\underline{r}}_{cl}^2\, d\tau \tag{D.7}$$

consider $\dot{\underline{r}}^2$. Substituting for one of the $\dot{\underline{r}}$ from (D.3) gives

$$m\dot{\underline{r}}^2 = \dot{\underline{r}}.(-\lambda\underline{n} + \underline{c}) \tag{D.8}$$

substituting for λ and for $\underline{n}.\dot{\underline{r}}$ from the constraint

$$m\dot{\underline{r}}^2 = \frac{(\underline{n}.\underline{c} + mn_t)n_t}{n^2} + \dot{\underline{r}}.\underline{c} \tag{D.9}$$

so

$$\begin{aligned} S_{cl} &= \frac{1}{2}\left(\int \frac{n_t}{n^2}\underline{n}\, d\tau + \Delta\underline{r}\right).\underline{c} + \frac{m}{2}\int \frac{n_t^2}{n^2} d\tau \\ &= \frac{m}{2}(\Delta\underline{r} + \Delta\underline{r}_0)^T M^{-1}(\Delta\underline{r} + \Delta\underline{r}_0) + \frac{m}{2}\int \frac{n_t^2}{n^2}\, d\tau \end{aligned} \tag{D.10}$$

where

$$\Delta\underline{r}_0 = \int_0^t \frac{n_t}{n^2}\underline{n}\, d\tau \tag{D.11}$$

and M is the matrix

$$\int_0^t \left(1 - \frac{1}{n^2}(\underline{n}\,\underline{n})\right) d\tau \tag{D.12}$$

D.3 Quantum

In order to evaluate the path integral

$$\int_{\underline{r}(0),0}^{\underline{r}(t),t} e^{\frac{i}{\hbar}\int_{\tau=0}^{t}(\frac{m}{2}\dot{\underline{r}}^2 + \lambda(\underline{n}.\dot{\underline{r}} + n_t))d\tau}\, d^\infty\underline{r}(\tau)\, d^\infty\lambda(\tau)$$

to obtain the propagator $K(\underline{r}_t, t; \underline{r}_0, 0)$, an integral is introduced to change the fixed end point path integral into one with free end points.

$$K = \int_0^t \int e^{\frac{i}{\hbar}\int_{\tau=0}^{t}(\frac{m}{2}\dot{\underline{r}}^2+\lambda(\underline{n}.\dot{\underline{r}}+n_t)+\hbar\underline{b}.\dot{\underline{r}})d\tau} e^{-i\underline{b}.\Delta\underline{r}}\, d^2\underline{b}\, d^\infty\dot{\underline{r}}(\tau)\, d^\infty\lambda(\tau) \tag{D.13}$$

the $\dot{\underline{r}}(\tau)$ path integral is performed first

$$K = \int\int e^{-\frac{i}{2\hbar m}\int_{\tau=0}^{t}(\lambda^2\underline{n}^2+2\lambda(\hbar\underline{b}.\underline{n}-mn_t)+\hbar^2\underline{b}^2)d\tau} e^{-i\underline{b}.\Delta\underline{r}}\, d^2\underline{b}\, d^\infty\lambda(\tau) \tag{D.14}$$

and then the $\lambda(\tau)$ path integral.

$$K = \int e^{-\frac{i\hbar}{2m}\int\left(\underline{b}^2-\frac{(\underline{b}.\underline{n})^2}{n^2}\right)d\tau - i\underline{b}.\left(\Delta\underline{r}+\int\frac{n_t}{n^2}\underline{n}\,d\tau\right)+\frac{i}{2\hbar}\int\frac{n_t^2}{n^2}d\tau} d^2\underline{b} \tag{D.15}$$

where $n = |\underline{n}|$

and finally the ordinary integral

$$K = \frac{2\pi m}{i\hbar\sqrt{\det(M)}} \exp\left(\frac{im}{2\hbar}(\Delta\underline{r}+\Delta\underline{r}_0)^T M^{-1}(\Delta\underline{r}+\Delta\underline{r}_0) + \frac{im}{2\hbar}\int_0^t \frac{n_t^2}{n^2}\,d\tau\right) \tag{D.16}$$

with the same definitions of $\Delta\underline{r}_0$ and M as in the preceeding section. Since the exponent of the original path integral is a quadratic form, the exponent in the result for K can be checked by verifying that it is $\frac{i}{\hbar}S_{cl}$.

Appendix E

Evaluation of the integral over a rhombus unit cell

$$I = \int_{\Delta x'=-\frac{a}{2}}^{\frac{a}{2}} \int_{y'=L_-(\Delta x')}^{L_+(\Delta x')} \cos\left(\frac{N\pi}{a}\Delta u'\right) \cos\left(\frac{n\pi}{a}\Delta x'\right) \exp(i(k_v v' - k_y y'))dy'd(\Delta x') \quad \text{(E.1)}$$

where $L_\pm(\Delta x') = y'_c \pm \frac{a}{2}\csc\Delta\theta - \Delta x'\cot\Delta\theta$

$\Delta x' = x' - x'_c$

$\Delta u' = u' - u'_c$

taking $\theta_{initial} = 0$ so that $\Delta\theta = \theta$

using

$$\begin{aligned} u' &= x'\cos\theta + y'\sin\theta \\ v' &= y'\cos\theta - x'\sin\theta \\ u'_c &= x'_c\cos\theta + y'_c\sin\theta \\ v'_c &= y'_c\cos\theta - x'_c\sin\theta \end{aligned}$$

and defining γ and Δ by:

$$y' = y'_c + a\gamma\csc\theta - (x' - x'_c)\cot\theta \quad \text{(E.2)}$$

$$\Delta = \frac{x' - x'_c}{a} \quad \text{(E.3)}$$

gives

$$I = \int_{\Delta=-\frac{1}{2}}^{\frac{1}{2}} \int_{\gamma=-\frac{1}{2}}^{\frac{1}{2}} \cos(n\pi\Delta)\cos(N\pi\gamma)e^{i(Aa\Delta + Ba\gamma + \lambda)}\left(\frac{a}{\sin\theta}\right)d\gamma\, a\, d\Delta \quad \text{(E.4)}$$

where

$A = k_y \cot\theta - k_v \csc\theta$

$B = k_v \cot\theta - k_y \csc\theta$

$\lambda = k_y(x'_c \cot\theta - u'_c \csc\theta) + k_v(u'_c \cot\theta - x'_c \csc\theta)$

after substituting for y'_c in terms of u'_c and x'_c.

So the double integral separates

$$I = \left(\int_{\Delta=-\frac{1}{2}}^{\frac{1}{2}} \cos(n\pi\Delta) e^{iAa\Delta} d\Delta \right) \left(\int_{\gamma=-\frac{1}{2}}^{\frac{1}{2}} \cos(N\pi\gamma) e^{iBa\gamma} d\gamma \right) e^{i\lambda} \left(\frac{a^2}{\sin\theta} \right) \tag{E.5}$$

Using the result

$$\int_{x=-\frac{1}{2}}^{\frac{1}{2}} \cos(mx)\, e^{ikx}\, dx = \frac{\sin(\frac{1}{2}(k+m))}{k+m} + \frac{(\sin\frac{1}{2}(k-m))}{k-m} \tag{E.6}$$

gives

$$\begin{aligned} I \;=\; & \left(\frac{a^2}{\sin\theta} \right) e^{i(x'_c A + u'_c B)} \times \\ & \left(\frac{\sin\frac{1}{2}(Aa+n\pi)}{Aa+n\pi} + \frac{\sin\frac{1}{2}(Aa-n\pi)}{Aa-n\pi} \right) \left(\frac{\sin\frac{1}{2}(Ba+N\pi)}{Ba+N\pi} + \frac{\sin\frac{1}{2}(Ba-N\pi)}{Ba-N\pi} \right) \end{aligned}$$

for n,N both odd this is

$$I = \frac{a^2}{\sin\Delta\theta} e^{i(j_1 aA + j_2 aB)} \frac{4\pi^2 nN}{[(Aa)^2 - (n\pi)^2][(Ba)^2 - (N\pi)^2]} \left(\cos\frac{Aa}{2} \cos\frac{Ba}{2} \right) (-1)^{\frac{1}{2}(n+N-2)} \tag{E.7}$$

where $j_1 = \frac{x'_c}{a}$, $j_2 = \frac{u'_c}{a}$.

Appendix F

The link between sum over images and modes for a single stage in 1D

To compare the 1D sum over images expression for the propagator with a 1D modes expression it is probably simplest to set the origin midway between the pair of planes (at the beginning of the calculation).

The expression for the propagator based on a sum over images is

$$K = \sqrt{\frac{\nu}{\pi i}} \left(\sum_{n=-\infty}^{\infty} e^{i\nu(\Delta x + 2an)^2} - \sum_{N=-\infty}^{\infty} e^{i\nu(\Delta x + 2\beta + 2aN)^2} \right) \tag{F.1}$$

where $\Delta x = x_{final} - x_{initial}$

$\nu = \frac{m}{2\hbar\epsilon}$ (mechanics)

or $\nu = \frac{k}{2\epsilon c}$ (optics)

ϵ = duration of stage in time

m = mass of particle

a is the "lane width"

β is the initial distance to the plane "below" the particle, i.e. the nearest plane in the direction of x decreasing.

So, with the current choice of origin for x, $\beta = x_{initial} + \frac{a}{2}$. This means that $\Delta x + 2\beta = x_f + x_i + a$.

Applying the Poisson summation formula to sums like those appearing in F.1

$$\sum_{n=-\infty}^{\infty} e^{i\nu(X+\alpha n)^2} = \frac{1}{\alpha} \sum_{m=-\infty}^{\infty} e^{-\frac{2i\pi m}{\alpha} X - \frac{i}{\nu}(\frac{\pi m}{\alpha})^2} \tag{F.2}$$

so, with $\alpha = 2a$

$$K = \frac{1}{2a} \sum_{m=-\infty}^{\infty} e^{-\frac{i}{4\nu}(\frac{\pi}{a})^2 m^2} \left[e^{-\frac{im\pi}{a}(x_f - x_i)} - e^{-\frac{im\pi}{a}(x_f + x_i + a)} \right] \tag{F.3}$$

the $m = 0$ term is zero so this may be re-written as

$$K = \frac{1}{2a}\sum_{m=1}^{\infty} e^{-\frac{i}{4\nu}(\frac{\pi}{a})^2 m^2}\left[e^{-\frac{im\pi}{a}x_f}\left(e^{\frac{im\pi}{a}x_i} - e^{-im\pi}e^{-\frac{im\pi}{a}x_i}\right)\right.$$
$$\left. + e^{\frac{im\pi}{a}x_f}\left(e^{-\frac{im\pi}{a}x_i} - e^{im\pi}e^{\frac{im\pi}{a}x_i}\right)\right] \qquad \text{(F.4)}$$

i.e.

$$K_{modes} = \frac{2}{a}\sum_{m\overset{odd}{=}1}^{\infty} e^{-\frac{i}{4\nu}(\frac{\pi}{a})^2 m^2}\cos\left(\frac{m\pi}{a}x_f\right)\cos\left(\frac{m\pi}{a}x_i\right)$$
$$+\frac{2}{a}\sum_{m\overset{even}{=}1}^{\infty} e^{-\frac{i}{4\nu}(\frac{\pi}{a})^2 m^2}\sin\left(\frac{m\pi}{a}x_f\right)\sin\left(\frac{m\pi}{a}x_i\right) \qquad \text{(F.5)}$$

which is invariant under the interchange of x_f and x_i. For brevity modes calculations are generally carried out using the first half of this result only, i.e. only the part with cosines and $m = \textit{odd integer}$ is considered explicitly (the presence of the second part being "understood").

The fact that the $m = 0$ term in the "sum over modes" is zero accords with expectations since this term represents the "constant mode" which does not satisfy the constraint requirement to be zero on the "constraint planes".

The version of the Poisson summation formula [28] (which links a sum of a function with the sum of its Fourier transform) required to obtain equation (F.2) is more general than the version given in the main text (equation (9.33)) i.e.

$$\sum_{n=-\infty}^{\infty} f(\alpha n) = \frac{1}{\alpha}\sum_{m=-\infty}^{\infty}\int_{-\infty}^{\infty} f(x)e^{i(\frac{2\pi m}{\alpha})x}dx \qquad \text{(F.6)}$$

This reduces to equation (9.33) when $\alpha = 1$. There are various ways to obtain the result (F.6): for example, a possible starting point is to consider the convolution of a function with a "delta comb".

Appendix G

Implementation of "phase screens" (for a single stage)

Before comparing "phase screens" with "modes" (appendix H), it is useful to introduce a modified method of implementing "phase screens" (method B). It is necessary to show that this is equivalent to the "direct" approach (method A) which may be summarised as:

A1 — take an incident "long sine" wave

A2 — "alternate" it

A3 — evaluate the diffraction (i.e. integrate over sources)

A5 — "unalternate"

Step A1 refers to a wavefunction with transverse dependence $\sin \frac{2\pi}{p_1 a}(x + a_1)$, $p_1 \gg 1$
Step A2 is passage through the phase screen at the beginning of the stage.
"Alternate" means multiply by the (odd) square-wave $f_{sign}(x, a)$ with Fourier series
$\frac{4}{\pi} \sum_{m=1}^{\infty}{}^{odd} \frac{1}{m} \sin\left(\frac{m\pi}{a}x\right)$
Step A3 is propagation through the "free space" between the phase screens.
Step A5 is passage through the phase screen at the end of the stage.
"Unalternate" means multiply by $f_{sign}(x, a)$.

Method B is:

B1 — take an incident long sine (same as A1)

B2 — "alternate" it (same as A2)

B3 — express the resulting function as a Fourier series

B4 — propagate the Fourier components

B5 — "unalternate" (same as A5)

The final steps are the same in both cases so need not be considered for a comparison. Since method B uses a Fourier series it is advantageous to introduce a step similar to B3 into method A, i.e.

A4 — express the resulting function as a Fourier series

To compare the methods it is sufficient to compare the "propagated Fourier components" of step B4 (B_c, B_s) with the Fourier components of step A4 (A_c, A_s).
where

$$\begin{Bmatrix} A_c \\ A_s \end{Bmatrix} = \frac{2}{p_1 a} \int_{x=-\frac{1}{2}p_1 a}^{\frac{1}{2}p_1 a} \begin{Bmatrix} \cos \\ \sin \end{Bmatrix} \frac{2\pi n x}{p_1 a} \int_{\mu=-\infty}^{\infty} f_{sign}(\mu) \sin\left(\frac{2\pi}{p_1 a}(\mu + \alpha_1)\right) e^{\frac{ik}{2z}(x-\mu)^2} d\mu\, dx \quad \text{(G.1)}$$

$$\begin{Bmatrix} B_c \\ B_s \end{Bmatrix} = \left(\frac{2}{p_1 a} \int_{y=-\frac{1}{2}p_1 a}^{\frac{1}{2}p_1 a} f_{sign}(y) \sin\left(\frac{2\pi}{p_1 a}(y + \alpha_1)\right) \begin{Bmatrix} \cos \\ \sin \end{Bmatrix} \left(\frac{2\pi n}{p_1 a} y\right) dy \right) \sqrt{\frac{2\pi i z}{k}} e^{-\frac{iz}{2k}\left(\frac{2n\pi}{p_1 a}\right)^2} \quad \text{(G.2)}$$

In fact the Fourier sum for the f_{sign} function can be "factored out" from these expressions and the comparison made between

$$\begin{Bmatrix} a_c \\ a_s \end{Bmatrix} = \int_{x=-\frac{1}{2}p_1 a}^{\frac{1}{2}p_1 a} \begin{Bmatrix} \cos \\ \sin \end{Bmatrix} (k_1 n x) \int_{\mu=-\infty}^{\infty} \sin\left(\frac{m\pi}{a}\mu\right) \sin\left(k_1(\mu + \alpha_1)\right) \sqrt{\frac{k}{2\pi i z}} e^{\frac{ik}{2z}(x-\mu)^2} d\mu\, dx \quad \text{(G.3)}$$

$$\begin{Bmatrix} b_c \\ b_s \end{Bmatrix} = \left(\int_{y=-\frac{1}{2}p_1 a}^{\frac{1}{2}p_1 a} \sin\left(\frac{m\pi}{a} y\right) \sin\left(k_1(y + \alpha_1)\right) \begin{Bmatrix} \cos \\ \sin \end{Bmatrix} k_1 n y\, dy \right) e^{-\frac{iz}{2k}(k_1 n)^2} \quad \text{(G.4)}$$

where $k_1 = \frac{2\pi}{p_1 a}$
Using the identity

$$\begin{aligned} \int_{\mu=-\infty}^{\infty} f(\mu) d\mu &= \sum_{l=-\infty}^{\infty} \int_{\mu=-\frac{1}{2}p_1 a + l p_1 a}^{\frac{1}{2}p_1 a + l p_1 a} f(\mu) d\mu \\ &= \sum_{l=-\infty}^{\infty} \int_{w=-\frac{1}{2}p_1 a}^{\frac{1}{2}p_1 a} f(w + l p_1 a) dw \end{aligned}$$

where $w = \mu - lp_1 a$

with

$$f(\mu) = \sin\left(\frac{m\pi}{a}\mu\right) \sin\left(k_1(\mu + \alpha_1)\right) e^{\frac{ik}{2z}(x-\mu)^2}$$

gives

$$\int_{\mu=-\infty}^{\infty} f(\mu)d\mu = \sum_{l=-\infty}^{\infty} \int_{w=-\frac{1}{2}p_1 a}^{\frac{1}{2}p_1 a} \sin\left(\frac{m\pi}{a}w\right) \sin\left(k_1(w + \alpha_1)\right) e^{\frac{ik}{2z}(w-x+lp_1 a)^2} dw \qquad \text{(G.5)}$$

where use has been made of

$$\begin{aligned} &\sin\left(\frac{m\pi}{a}(w + lp_1 a)\right) \sin\left(k_1(w + lp_1 a + \alpha_1)\right) \\ &= \left(\sin\left(\frac{m\pi}{a}w\right) \cos(mlp_1\pi)\right) \left(\sin\left(\frac{2\pi}{p_1 a}(a + \alpha_1)\right) \cos(2\pi l)\right) \\ &= \sin\left(\frac{m\pi}{a}w\right) \sin\left(k_1(w + \alpha_1)\right) \end{aligned} \qquad \text{(G.6)}$$

since l, m, p_1 are all integers and p_1 is taken to be even.

Using the Poisson summation formula on the sum over l gives

$$\sum_{l=-\infty}^{\infty} e^{\frac{ik}{2z}(w-x+lp_1 a)^2} = \frac{1}{p_1 a} \sum_{N=-\infty}^{\infty} \sqrt{\frac{2\pi i z}{k}}\, e^{-\frac{iz}{2k}\left(\frac{2N\pi}{p_1 a}\right)^2} e^{-i\left(\frac{2N\pi}{p_1 a}(w-x)\right)} \qquad \text{(G.7)}$$

So

$$\begin{aligned} \begin{Bmatrix} a_c \\ a_s \end{Bmatrix} &= \frac{1}{p_1 a} \int_{w=-\frac{1}{2}p_1 a}^{\frac{1}{2}p_1 a} \sin\left(\frac{m\pi}{a}w\right) \sin\left(k_1(w + \alpha_1)\right) \\ &\quad \times \sum_{N=-\infty}^{\infty} e^{-\frac{iz}{2k}(k_1 N)^2} \int_{x=-\frac{1}{2}p_1 a}^{\frac{1}{2}p_1 a} \begin{Bmatrix} \cos \\ \sin \end{Bmatrix}(k_1 nx) e^{-ik_1 N(w-x)}\, dx\, dw \end{aligned} \qquad \text{(G.8)}$$

$$\begin{aligned} \begin{Bmatrix} g_c \\ g_s \end{Bmatrix}(w, n, N) &\equiv \int_{x=-\frac{1}{2}p_1 a}^{\frac{1}{2}p_1 a} \begin{Bmatrix} \cos \\ \sin \end{Bmatrix}(k_1 nx) e^{-ik_1 N(w-x)}\, dx \\ &= \frac{p_1 a}{2} e^{-ik_1 Nw} \left[\begin{Bmatrix} +1 \\ -i \end{Bmatrix} \mathrm{sinc}\left((N + n)\pi\right) + \begin{Bmatrix} +1 \\ +i \end{Bmatrix} \mathrm{sinc}\left((N - n)\pi\right) \right] \end{aligned} \qquad \text{(G.9)}$$

N and n are integers, so

$$\sum_{N=-\infty}^{\infty} e^{-\frac{iz}{2k}(k_1 N)^2} \begin{Bmatrix} g_c \\ g_s \end{Bmatrix}(w, n, N) = p_1 a \begin{Bmatrix} \cos \\ \sin \end{Bmatrix}(k_1 nw) e^{-\frac{iz}{2k}(k_1 n)^2} \qquad \text{(G.10)}$$

Consequently

$$\begin{Bmatrix} a_c \\ a_s \end{Bmatrix} = \left(\int_{w=-\frac{1}{2}p_1 a}^{\frac{1}{2}p_1 a} \sin\left(\frac{m\pi}{a}w\right) \sin\left(k_1(w+\alpha_1)\right) \begin{Bmatrix} \cos \\ \sin \end{Bmatrix} (k_1 n w)\, dw \right) e^{-\frac{iz}{2k}(k_1 n)^2} \qquad \text{(G.11)}$$

So $\begin{Bmatrix} a_c \\ a_s \end{Bmatrix} = \begin{Bmatrix} b_c \\ b_s \end{Bmatrix}$, as required for methods A and B to be equivalent. For sake of completeness, the integral can be evaluated to give

$$\begin{aligned} &\int_{w=-\frac{1}{2}p_1 a}^{\frac{1}{2}p_1 a} \sin\left(\frac{m\pi}{a}w\right) \sin\left(k_1(w+\alpha_1)\right) \begin{Bmatrix} \cos \\ \sin \end{Bmatrix} (k_1 n w)\, dw \\ &= \frac{p_1 a}{4} \left[\operatorname{sinc}\left(\pi\left(1 - \frac{1}{2}mp_1 + n\right)\right) \begin{Bmatrix} \cos \\ \sin \end{Bmatrix} (k_1\alpha_1) \right. \\ &\quad + \operatorname{sinc}\left(\pi\left(1 - \frac{1}{2}mp_1 - n\right)\right) \begin{Bmatrix} \cos \\ -\sin \end{Bmatrix} (k_1\alpha_1) \\ &\quad + \operatorname{sinc}\left(\pi\left(1 + \frac{1}{2}mp_1 + n\right)\right) \begin{Bmatrix} -\cos \\ -\sin \end{Bmatrix} (k_1\alpha_1) \\ &\quad \left. + \operatorname{sinc}\left(\pi\left(1 + \frac{1}{2}mp_1 - n\right)\right) \begin{Bmatrix} -\cos \\ \sin \end{Bmatrix} (k_1\alpha_1) \right] \end{aligned} \qquad \text{(G.12)}$$

Appendix H

Comparison of "phase screens" and "modes"

H.1 Modes

Consider the long sine wave

$$\sin\left(k_1(x+\alpha_1)\right)$$

where $k_1 = \frac{2\pi}{p_1 a}$

$p_1 \gg 1$, p_1 is an even integer

Define $T(x;a)$ to be a "Top-hat function" height 1, width a,

i.e. $T(y;a) = 1$ for $|y| < \frac{a}{2}$, $T(y;a) = 0$ for $|y| > \frac{a}{2}$

So the function $g(x;j,a) \equiv T(x-(j+\frac{1}{2})a;a)\sin\left(k_1(x+\alpha_1)\right)$ is the "chunk" of the sine between $x = ja$ and $x = (j+1)a$, it is zero outside this interval.

Take this function and shift it by ja to move its non-zero part into the first "lane" (i.e. substitute $x = y + ja$)

Define a new function to be identical to this for $y > 0$ and odd i.e. $f(-y) = -f(y)$

Find the Fourier series for this function on the interval $[-a,a]$ i.e.

$$F_1(y) = \sum_{n=1}^{\infty} b_n(j,p_1,\alpha_1)\sin\left(\frac{n\pi}{a}y\right) \tag{II.1}$$

$$b_n(j,p_1,\alpha_1) = \frac{2}{a}\int_{Y=0}^{a} \sin\left(k_1(Y+ja+\alpha_1)\right)\sin\left(\frac{n\pi}{a}Y\right)dY \tag{II.2}$$

Shift this back again by the same amount, i.e. substitute $y = x - ja$ in $F_1(y)$

The resulting Fourier series may be written

$$f_1(x) = \sum_{n=1}^{\infty} B_n(j, p_1, \alpha_1) \sin\left(\frac{n\pi}{a}x\right) \tag{H.3}$$

where

$$B_n(j, p_1, \alpha_1) = (-1)^{nj} b_n(j, p_1, \alpha_1) \tag{H.4}$$

Propagate the modes by multiplying each by $\exp\left(-\frac{iz}{2k}\left(\frac{n\pi}{a}\right)^2\right)$

Multiply by $T(x-(j+\frac{1}{2})a;a)$ to remove everything outside the interval $x \in [ja,(j+1)a]$

Then sum over j to include contributions from all "chunks" of the original sine wave, i.e.

$$f_p(x) \equiv \sum_{j=-\infty}^{\infty} T\left(x-\left(j+\frac{1}{2}\right)a;a\right) \sum_{n=1}^{\infty} B_n(j,p_1,\alpha_1) \sin\frac{n\pi x}{a} \exp\left(-\frac{iz}{2k}\left(\frac{n\pi}{a}\right)^2\right) \tag{H.5}$$

Consider the overlap with another long sine-wave

$$I_{OL} = \frac{1}{2La} \int_{x=-La}^{La} \sin\left(k_2(x+\alpha_2)\right) f_p(x)\, dx \tag{H.6}$$

where $k_1 = \frac{2\pi}{p_1 a}$

$p_2 \gg 1$, p_2 is an even integer

Defining $\gamma = -\frac{z}{2k}\left(\frac{\pi}{a}\right)^2$

$$\begin{aligned} I_{OL} &= \frac{1}{2La} \int_{x=-La}^{La} \sin\left(k_2(x+\alpha_2)\right) \\ &\quad \times \left[\sum_{j=-\infty}^{\infty} T\left(x-\left(j+\frac{1}{2}\right)a;a\right) \sum_{n=1}^{\infty} B_n(j,p_1,\alpha_1) \sin\frac{n\pi x}{a} e^{i\gamma n^2}\right] dx \\ &= \frac{1}{4L} \sum_{j=-L}^{L-1} T\left(x-\left(j+\frac{1}{2}\right)a;a\right) \\ &\quad \times \sum_{n=1}^{\infty} \left[\frac{2}{a} \int_{x=ja}^{(j+1)a} \sin\left(k_2(x+\alpha_2)\right) \sin\frac{n\pi x}{a} dx\right] B_n(j,p_1,\alpha_1) e^{i\gamma n^2} \end{aligned} \tag{H.7}$$

the substitution $u = x - ja$ shows that

$$\frac{2}{a} \int_{x=ja}^{(j+1)a} \sin\left(k_2(x+\alpha_2)\right) \sin\frac{n\pi x}{a} dx = B_n(j,p_2,\alpha_2) \tag{H.8}$$

and since

$$B_n(j,p_1,\alpha_1) B_n(j,p_2,\alpha_2) = b_n(j,p_1,\alpha_1) b_n(j,p_2,\alpha_2) \tag{H.9}$$

Equation (H.7) becomes

$$I_{OL} = \frac{1}{4L} \sum_{j=-L}^{L-1} T\left(x-\left(j+\frac{1}{2}\right)a;a\right) \sum_{n=1}^{\infty} b_n(j,p_1,\alpha_1) b_n(j,p_2,\alpha_2) e^{i\gamma n^2} \tag{H.10}$$

where

$$b_n(j,p,\alpha) = \frac{4}{\pi}\frac{n}{n^2-\left(\frac{2}{p}\right)^2}\left\{\begin{matrix}-\cos\\ \sin\end{matrix}\right\}\left(\frac{2\pi}{pa}\left(ja+\alpha+\frac{a}{2}\right)\right)\left\{\begin{matrix}\sin\\ \cos\end{matrix}\right\}\frac{\pi}{p} \qquad \text{(H.11)}$$

for n $\left\{\begin{matrix}even\\ odd\end{matrix}\right\}$

so the j dependent part of $b_n(j,p_1,\alpha_1)b_n(j,p_2,\alpha_2)$ is

$$\left\{\begin{matrix}\cos\\ \sin\end{matrix}\right\}(A_1j+B_1)\left\{\begin{matrix}\cos\\ \sin\end{matrix}\right\}(A_2j+B_2) = \frac{1}{2}\left((-1)^n\cos[(A_1+A_2)j+(B_1+B_2)]\right.$$
$$\left.+\cos[(A_1-A_2)j+(B_1-B_2)]\right) \qquad \text{(H.12)}$$

where $A_i = \frac{2\pi}{p_i}$, $B_i = \frac{2\pi}{p_i a}(\alpha_i + \frac{a}{2})$

So, to perform the sum over j first, it is necessary to evaluate sums of the form $\sum_{j=-L}^{L-1}\cos(aj+b)$. This may be achieved using the standard result

$$\sum_{k=0}^{n}\cos(ka+b) = \frac{\sin(\frac{n+1}{2}a)\cos(\frac{n}{2}a+b)}{\sin\frac{a}{2}} \qquad \text{(H.13)}$$

to show that

$$\sum_{j=-L}^{L-1}\cos(aj+b) = \frac{\sin(La)\cos(b-\frac{a}{2})}{\sin\frac{a}{2}} \qquad \text{(H.14)}$$

So

$$\sum_{j=-L}^{L-1}\frac{1}{2}\left((-1)^n\cos[(A_1+A_2)j+(B_1+B_2)]+\cos[(A_1-A_2)j+(B_1-B_2)]\right)$$
$$= \frac{1}{2}\left((-1)^n\frac{\sin\left(2L\left(\frac{1}{p_1}+\frac{1}{p_2}\right)\pi\right)\cos(\beta_1+\beta_2)}{\sin\left(\left(\frac{1}{p_1}+\frac{1}{p_2}\right)\pi\right)}+\frac{\sin\left(2L\left(\frac{1}{p_1}-\frac{1}{p_2}\right)\pi\right)\cos(\beta_1+\beta_2)}{\sin\left(\left(\frac{1}{p_1}-\frac{1}{p_2}\right)\pi\right)}\right) \qquad \text{(H.15)}$$

where $\beta_i = \frac{2\pi}{p_i a}\alpha_i$

So

$$I_{OL} = \frac{2}{\pi^2 L}\sum_{n=1}^{\infty}\frac{n^2}{[n^2-(\frac{2}{p_1})^2][n^2-(\frac{2}{p_2})^2]}\left\{\begin{matrix}F_{even}\\ F_{odd}\end{matrix}\right\}\exp\left(-\frac{iz}{2k}\left(\frac{n\pi}{a}\right)^2\right) \qquad \text{(H.16)}$$

where

$$\left\{\begin{matrix}F_{even}\\ F_{odd}\end{matrix}\right\} = \left(\frac{\sin\left(2L\left(\frac{1}{p_1}-\frac{1}{p_2}\right)\pi\right)\cos(\beta_1+\beta_2)}{\sin\left(\left(\frac{1}{p_1}-\frac{1}{p_2}\right)\pi\right)}\right.$$
$$\left.+(-1)^n\frac{\sin\left(2L\left(\frac{1}{p_1}+\frac{1}{p_2}\right)\pi\right)\cos(\beta_1+\beta_2)}{\sin\left(\left(\frac{1}{p_1}+\frac{1}{p_2}\right)\pi\right)}\right)$$
$$\times\left(\left\{\begin{matrix}\sin\\ \cos\end{matrix}\right\}\frac{\pi}{p_1}\right)\left(\left\{\begin{matrix}\sin\\ \cos\end{matrix}\right\}\frac{\pi}{p_2}\right) \qquad \text{(H.17)}$$

for n $\left\{\begin{matrix}even\\ odd\end{matrix}\right\}$

where $2L = p_2$, i.e. the wavelength of the overlap sine-wave in units of a.

H.2 Phase screens

Consider the long sine wave

$$\sin(k_1(x+\alpha_1))$$

where $k_1 = \frac{2\pi}{p_1 a}$

$p_1 \gg 1$, p_1 is an even integer

"Alternate" it by multiplying by the periodic antisymmetric step function $f_{sign}(x)$, which has the Fourier series

$$f_{sign}(x) = \frac{4}{\pi} \sum_{m \overset{odd}{=} 1}^{\infty} \frac{1}{m} \sin\left(\frac{m\pi}{a}x\right) \tag{H.18}$$

Find the Fourier series of this "composite" function i.e.

$$\frac{1}{2}A_0 + \sum_{n=1}^{\infty} (A_n \cos(k_1 n x) + B_n \sin(k_1 n x))$$

where

$$\begin{Bmatrix} A_n \\ B_n \end{Bmatrix} = \frac{2}{p_1 a} \int_{v=-\frac{1}{2}p_1 a}^{\frac{1}{2}p_1 a} f_{sign}(v) \sin(k_1(v+\alpha_1)) \begin{Bmatrix} \cos \\ \sin \end{Bmatrix} (k_1 n v)\, dv$$

"Propagate" the Fourier components by multiplying each by $\exp\left(-\frac{iz}{2k}\left(\frac{2n\pi}{p_1 a}\right)^2\right)$

"Unalternate" by re-multiplying by $f_{sign}(x)$, to give the phase screens propagated function.

Consider the overlap with another long sine wave i.e.

$$I_{OL} = \frac{1}{4}a_0 A_0 + \frac{1}{2}\sum_{n=1}^{\infty}(a_n A_n + b_n B_n) \exp\left(i\gamma\left(\frac{2n}{p_1}\right)^2\right) \tag{H.19}$$

where

$$\begin{Bmatrix} a_n \\ b_n \end{Bmatrix} = \frac{2}{p_2 a} \int_{x=-\frac{1}{2}p_2 a}^{\frac{1}{2}p_2 a} f_{sign}(x) \sin(k_2(x+\alpha_2)) \begin{Bmatrix} \cos \\ \sin \end{Bmatrix} (k_1 n x)\, dx \tag{H.20}$$

(i.e. $L = \frac{1}{2}p_2$)

$$\gamma = -\frac{z}{2k}\left(\frac{\pi}{a}\right)^2 \tag{H.21}$$

and $p_2 \gg 1$ (p_2 is even)

Using the result

$$\begin{aligned} &\int_{x=-L}^{L} \sin(\lambda x + \beta) \sin(\mu x) \begin{Bmatrix} \cos \\ \sin \end{Bmatrix} (\nu x)\, dx \\ &= \frac{1}{2}\left(\frac{\sin((\nu+\lambda-\mu)L)}{\nu+\lambda-\mu}\begin{Bmatrix} \cos \\ \sin \end{Bmatrix}\beta + \frac{\sin((\nu-\lambda+\mu)L)}{\nu-\lambda+\mu}\begin{Bmatrix} \cos \\ -\sin \end{Bmatrix}\beta \right. \\ &\left. - \frac{\sin((\nu+\lambda+\mu)L)}{\nu+\lambda+\mu}\begin{Bmatrix} \cos \\ \sin \end{Bmatrix}\beta - \frac{\sin((\nu-\lambda-\mu)L)}{\nu-\lambda-\mu}\begin{Bmatrix} \cos \\ -\sin \end{Bmatrix}\beta \right) \end{aligned} \tag{H.22}$$

it may be shown that

$$a_n = \frac{2}{\pi}\sum_{m=1}^{\infty}{}^{odd}\frac{1}{m}\cos\beta_2\left[\mathrm{sinc}\left(\left(\frac{p_2}{p_1}n+1-m\frac{p_2}{2}\right)\pi\right)+\mathrm{sinc}\left(\left(\frac{p_2}{p_1}n-1+m\frac{p_2}{2}\right)\pi\right)\right.$$
$$\left.-\mathrm{sinc}\left(\left(\frac{p_2}{p_1}n+1+m\frac{p_2}{2}\right)\pi\right)-\mathrm{sinc}\left(\left(\frac{p_2}{p_1}n-1-m\frac{p_2}{2}\right)\pi\right)\right] \quad \text{(H.23)}$$

$$b_n = \frac{2}{\pi}\sum_{m=1}^{\infty}{}^{odd}\frac{1}{m}\sin\beta_2\left[\mathrm{sinc}\left(\left(\frac{p_2}{p_1}n+1-m\frac{p_2}{2}\right)\pi\right)-\mathrm{sinc}\left(\left(\frac{p_2}{p_1}n-1+m\frac{p_2}{2}\right)\pi\right)\right.$$
$$\left.-\mathrm{sinc}\left(\left(\frac{p_2}{p_1}n+1+m\frac{p_2}{2}\right)\pi\right)+\mathrm{sinc}\left(\left(\frac{p_2}{p_1}n-1-m\frac{p_2}{2}\right)\pi\right)\right] \quad \text{(H.24)}$$

$$A_n = \frac{2}{\pi}\sum_{M=1}^{\infty}{}^{odd}\frac{1}{M}\cos\beta_1\left[\mathrm{sinc}\left(\left(n+1-M\frac{p_1}{2}\right)\pi\right)+\mathrm{sinc}\left(\left(n-1+M\frac{p_1}{2}\right)\pi\right)\right.$$
$$\left.-\mathrm{sinc}\left(\left(n+1+M\frac{p_1}{2}\right)\pi\right)-\mathrm{sinc}\left(\left(n-1-M\frac{p_1}{2}\right)\pi\right)\right] \quad \text{(H.25)}$$

$$B_n = \frac{2}{\pi}\sum_{M=1}^{\infty}{}^{odd}\frac{1}{M}\sin\beta_1\left[\mathrm{sinc}\left(\left(n+1-M\frac{p_1}{2}\right)\pi\right)-\mathrm{sinc}\left(\left(n-1+M\frac{p_1}{2}\right)\pi\right)\right.$$
$$\left.-\mathrm{sinc}\left(\left(n+1+M\frac{p_1}{2}\right)\pi\right)+\mathrm{sinc}\left(\left(n-1-M\frac{p_1}{2}\right)\pi\right)\right] \quad \text{(H.26)}$$

where $\beta_i = \frac{2\pi}{p_1 a}\alpha_i$

(So $A_0 = 0$ and $B_0 = 0$)

Since n,M and $\frac{p_1}{2}$ are integers, the sinc functions in the expressions for A_n and B_n have arguments which are an integer multiple of π and so they may be written as delta functions. The restrictions $n \geq 1$, $M \geq 1$ and $\frac{p_1}{2} > 2$ ensure that only two of the four functions provide contributions. These contributions are for $n = m\frac{p_1}{2} \pm 1$. So

$$I_{OL} = \frac{1}{\pi}\sum_{M=1}^{\infty}{}^{odd}\cos\beta_1\left[a_{M\frac{p_1}{2}-1}\exp\left(i\gamma\left(M-\frac{2}{p_1}\right)^2\right)-a_{M\frac{p_1}{2}+1}\exp\left(i\gamma\left(M+\frac{2}{p_1}\right)^2\right)\right]$$
$$+\frac{1}{\pi}\sum_{M=1}^{\infty}{}^{odd}\sin\beta_1\left[b_{M\frac{p_1}{2}-1}\exp\left(i\gamma\left(M-\frac{2}{p_1}\right)^2\right)-b_{M\frac{p_1}{2}+1}\exp\left(i\gamma\left(M+\frac{2}{p_1}\right)^2\right)\right] \quad \text{(H.27)}$$

To obtain $a_{M\frac{p_1}{2}\pm 1}$ and $b_{M\frac{p_1}{2}\pm 1}$ substitute $n = M\frac{p_1}{2} \pm 1$ in the expressions for a_n and b_n. Simplification may be achieved by considering the sin x and $\frac{1}{x}$ parts of sinc x separately and using $\sin(N\pi+\theta) = (-1)^N \sin\theta$ for $N \in$ *integers*. The remaining sinc factors are common to all four terms. Hence they may be factored out and the $\frac{1}{x}$ terms collected in

pairs. Since m is always odd $(-1)^{m\frac{p_2}{2}} = (-1)^{\frac{p_2}{2}}$.

$$a_{M\frac{p_1}{2}\pm 1} = \frac{2}{\pi^2}\cos\beta_2 \sin\left(\pm\frac{p_2}{p_1}\pi\right)(-1)^{M\frac{p_2}{2}}(-1)^{\frac{p_2}{2}} \times \sum_{m\overset{odd}{=}1}^{\infty} -\frac{4}{p_2}\left[\frac{1}{\left(M\pm\frac{2}{p_1}-\frac{2}{p_2}\right)^2 - m^2} - \frac{1}{\left(M\pm\frac{2}{p_1}+\frac{2}{p_2}\right)^2 - m^2}\right] \quad \text{(H.28)}$$

$$b_{M\frac{p_1}{2}\pm 1} = \frac{2}{\pi^2}\sin\beta_2 \sin\left(\pm\frac{p_2}{p_1}\pi\right)(-1)^{M\frac{p_2}{2}}(-1)^{\frac{p_2}{2}} \times \sum_{m\overset{odd}{=}1}^{\infty} -\frac{4}{p_2}\left[\frac{1}{\left(M\pm\frac{2}{p_1}-\frac{2}{p_2}\right)^2 - m^2} + \frac{1}{\left(M\pm\frac{2}{p_1}+\frac{2}{p_2}\right)^2 - m^2}\right] \quad \text{(H.29)}$$

The standard result

$$S(x) = \sum_{k=1}^{\infty}\frac{1}{x^2-k^2} = \frac{\pi}{2x}\cot(\pi x) - \frac{1}{2x^2}$$

may be used to obtain

$$\sum_{k\overset{odd}{=}1}^{\infty}\frac{1}{x^2-k^2} = S(x) - \frac{1}{4}S\left(\frac{x}{2}\right) = -\frac{\pi}{4x}\tan\frac{\pi x}{2}$$

Also $\tan(\frac{1}{2}M\pi + \phi) = -\cot\phi$ for M an odd integer.

Using these results gives

$$a_{M\frac{p_1}{2}\pm 1} = \frac{1}{\pi p_2}\cos\beta_2 \sin\left(\pm\frac{p_2}{p_1}\pi\right) \times(-1)^{M\frac{p_2}{2}}(-1)^{\frac{p_2}{2}}\left[\frac{\cot\left(\left(\pm\frac{1}{p_1}-\frac{1}{p_2}\right)\pi\right)}{\frac{M}{2}\pm\frac{1}{p_1}-\frac{1}{p_2}} - \frac{\cot\left(\left(\pm\frac{1}{p_1}+\frac{1}{p_2}\right)\pi\right)}{\frac{M}{2}\pm\frac{1}{p_1}+\frac{1}{p_2}}\right] \quad \text{(H.30)}$$

$$b_{M\frac{p_1}{2}\pm 1} = \frac{1}{\pi p_2}\sin\beta_2 \sin\left(\pm\frac{p_2}{p_1}\pi\right) \times(-1)^{M\frac{p_2}{2}}(-1)^{\frac{p_2}{2}}\left[\frac{\cot\left(\left(\pm\frac{1}{p_1}-\frac{1}{p_2}\right)\pi\right)}{\frac{M}{2}\pm\frac{1}{p_1}-\frac{1}{p_2}} + \frac{\cot\left(\left(\pm\frac{1}{p_1}+\frac{1}{p_2}\right)\pi\right)}{\frac{M}{2}\pm\frac{1}{p_1}+\frac{1}{p_2}}\right] \quad \text{(H.31)}$$

Substituting from (H.30) and (H.31) into the expression (H.27) for the overlap and using $(-1)^{M\frac{p_2}{2}} = (-1)^{\frac{p_2}{2}}$ for odd integers M.

Then collecting together terms in $\cot\left(\left(\frac{1}{p_1}+\frac{1}{p_2}\right)\pi\right)$ and in $\cot\left(\left(\frac{1}{p_1}-\frac{1}{p_2}\right)\pi\right)$, combining those containing β_1 and β_2 to form double angle cosines and then rewriting the exponential parts in terms of cos and sin gives

$$I_{OL} = \frac{4}{p_2\pi^2}\sin\left(\frac{p_2}{p_1}\pi\right)e^{i\gamma\left(\frac{4}{p_1\pi^2}\right)}\sum_{M\overset{odd}{=}1}^{\infty}\frac{1}{M^2}e^{i\gamma M^2}G_M \quad \text{(H.32)}$$

where

$$G_M = \cot\left(\left(\frac{1}{p_1}+\frac{1}{p_2}\right)\pi\right)\cos(\beta_1-\beta_2)\left[\frac{\cos\frac{\gamma M}{p_1}-\frac{i}{M}\left(\frac{2}{p_1}+\frac{2}{p_2}\right)\sin\frac{\gamma M}{p_1}}{1-\frac{1}{M^2}\left(\frac{2}{p_1}+\frac{2}{p_2}\right)^2}\right]$$
$$-\cot\left(\left(\frac{1}{p_1}-\frac{1}{p_2}\right)\pi\right)\cos(\beta_1+\beta_2)\left[\frac{\cos\frac{\gamma M}{p_1}-\frac{i}{M}\left(\frac{2}{p_1}-\frac{2}{p_2}\right)\sin\frac{\gamma M}{p_1}}{1-\frac{1}{M^2}\left(\frac{2}{p_1}-\frac{2}{p_2}\right)^2}\right] \quad \text{(H.33)}$$

H.3 Comparison

Define

$$c_i = \cos\frac{\pi}{p_i}$$
$$s_i = \sin\frac{\pi}{p_i}$$
$$C_i = \cos\beta_i$$
$$S_i = \sin\beta_i$$
$$\beta_i = \frac{2\pi}{p_i a}\alpha_i$$

Then removing all factors which modes and phase screen expressions have in common leaves a comparison between:

For n odd

Modes:

$$\frac{1}{\left[1-(\frac{2}{np_1})^2\right]\left[1-(\frac{2}{np_2})^2\right]}\left(\frac{C_1C_2-S_1S_2}{\sin\left(\left(\frac{1}{p_1}+\frac{1}{p_2}\right)\pi\right)}c_1c_2-\frac{C_1C_2+S_1S_2}{\sin\left(\left(\frac{1}{p_1}-\frac{1}{p_2}\right)\pi\right)}c_1c_2\right)$$

Phase screens:

$$(c_1c_2-s_1s_2)(C_1C_2+S_1S_2)\csc\left(\left(\frac{1}{p_1}+\frac{1}{p_2}\right)\pi\right)\left[\frac{\cos\frac{\gamma M}{p_1}-\frac{i}{M}\left(\frac{2}{p_1}+\frac{2}{p_2}\right)\sin\frac{\gamma M}{p_1}}{1-\frac{1}{M^2}\left(\frac{2}{p_1}+\frac{2}{p_2}\right)^2}\right]$$
$$-(c_1c_2-s_1s_2)(C_1C_2-S_1S_2)\csc\left(\left(\frac{1}{p_1}-\frac{1}{p_2}\right)\pi\right)\left[\frac{\cos\frac{\gamma M}{p_1}-\frac{i}{M}\left(\frac{2}{p_1}-\frac{2}{p_2}\right)\sin\frac{\gamma M}{p_1}}{1-\frac{1}{M^2}\left(\frac{2}{p_1}-\frac{2}{p_2}\right)^2}\right]$$

So modes and phase screens agree until terms of order $\left(\frac{1}{p_ip_j}\right)$ for n odd (and odd M)

For n even

Modes:

$$\frac{1}{p_2}\frac{1}{\left[1-(\frac{2}{np_1})^2\right]\left[1-(\frac{2}{np_2})^2\right]}\left(-\frac{C_1C_2-S_1S_2}{\sin\left(\left(\frac{1}{p_1}+\frac{1}{p_2}\right)\pi\right)}-\frac{C_1C_2+S_1S_2}{\sin\left(\left(\frac{1}{p_1}-\frac{1}{p_2}\right)\pi\right)}\right)s_1s_2$$
$$=-\frac{1}{p_2}\frac{2s_1s_2}{(s_1c_2)^2-(s_2c_1)^2}(C_1C_2s_1c_2+S_1S_2s_2c_1)$$

Phase screens:

0 (M odd)

So modes and phase screens agree until terms of order $\left(\frac{1}{p_i p_j}\right)$ for even n also.

Appendix I

Notation

The type of notation listed here consists of those symbols which may appear without introduction because they are used throughout the report. Other quantities are explicitly defined when they are used and are not included in this list.

The notation is given, followed by an explanation or reference to its definition.

- n_x, n_y, n_t functions defining the constraint for a system with 2 space dimensions (in the optical analogy time, t, becomes the coordinate in the"paraxial direction" i.e. z). In the simple simple nonholonomic system considered from chapter 5 onwards $n_t = 0$. They are the components of a vector normal to the constraint planelet in 3D space-time (section 1.4).

- $\underline{n}$ a vector function of time (or z in the optical analogy) which defines the constraint in the simple nonholonomic system considered from chapter 5 onwards. It is often referred to as the "constraint normal vector" because $n_t = 0$ and $\underline{n}$ is normal to the projection of the constraint planelet in the x-y plane.

- a the width of a "lane" as introduced in section 6.2.

- $\bar{a}$ the "average" width of a "lane" (used when a distribution of "lane widths" are considered) as introduced in section 9.2.2.

- $\hbar$ takes its conventional definition as Plank's constant divided by 2π

- c the speed of light in a vacuum (it is used in section 6.3 for example).

- ϵ = duration of a single stage (i.e. its length on the time axis)

- $\delta z = c\epsilon$ length of a single stage in the z-direction.

- $\Delta x = x_{final} - x_{initial}$ displacement in the x-direction.
- m = mass of particle (e.g. section 6.3)
- μ is defined in equation (8.12) for mechanics and equation (8.11) for optics
- ν is defined in equation (6.2) for mechanics and equation (6.3) for optics
- $\underline{r}$ position vector
- q^i i^{th} component of generalised coordinate vector

Bibliography

[1] **M.Abramowitz and I.A.Stegun (Eds.)** *Handbook of Mathematical functions* (Dover, New York, 1965)

[2] **V.I.Arnol'd (Ed.)** *Dynamical systems III* (Springer, Berlin, 1988)

[3] **V.I.Arnol'd** *Geometrical methods in the theory of ordinary differential equations* (Springer, New York, 1983)

[4] **C.M.Bender and S.A.Orszag** *Advanced Mathematical Methods for Scientists and Engineers* (McGraw-Hill, Singapore, 1978)

[5] **M.V.Berry** *Proc. R. Soc. Lond. A* **422** (1989) 7–21

[6] **M.V.Berry and J.Goldberg** *Nonlinearity* **1** (1988) 1–26

[7] **M.L.Boas** *Mathematical Methods in the Physical Sciences* (Wiley, New York, 1966)

[8] **E.A.Desloge** *Classical mechanics* (Wiley, New York, 1982)

[9] **P.A.M.Dirac** *Canad. J. Math.* **2** (1950) 129–148

[10] **R.P.Feynman and A.R.Hibbs** *Quantum mechanics and Path integrals* (McGraw-Hill, New York, 1965)

[11] **R.P.Feynman** *Revs. Modern Phys.* **20** (1948) 367–387

[12] **H.Goldstein** *Classical mechanics* (Addison-Wesley, Massachusetts, 1950)

[13] **G.R.Grimmett and D.R.Stirzaker** *Probability and Random Processes* (Oxford University Press, Oxford, 1982)

[14] **M.C.Gutzwiller** *Chaos in Classical and Quantum Mechanics* (Springer-Verlag, New York, 1990)

[15] **J.H.Hannay**, Ph.D. thesis, University of Cambridge, 1977 (unpublished)

[16] **J.H.Hannay, J.P.Keating and A.M.Ozorio de Almeida** *Nonlinearity* **7** (1994) 1327–1342

[17] **H.Hertz** *The principles of mechanics presented in a new form* (Dover, 1956)

[18] **H.Jeffreys** *Quart. J. Mech. and Appl. Math.* **7** (1954) 335–337

[19] **H.Jensen and H.Koppe** *Ann. of Phys.* **63** (1971) 586–591

[20] **F.Klein** *Elementary mathematics from an advanced standpoint (vol. 2)* (New York, 1939)

[21] **G.A.Korn and T.M.Korn** *Mathematical handbook for Scientists and Engineers* (McGraw-Hill, New York, 1961)

[22] **C.Lanczos** *The variational principles of mechanics* (University of Toronto press, Toronto, 1949)

[23] **A.D.Lewis and R.M.Murray** *Int. J. Non-Linear Mech.* **30** 793–815 (1995)

[24] **P.Maraner** *J. Phys. A* **28** (1995) 2939–2951

[25] **R.Martin**, 3rd year project (T16), University of Bristol, 1989 (unpublished)

[26] **J.Mathews and R.L.Walker** *Mathematical Methods of Physics* (Benjamin, New York, 1970)

[27] **I.Mitsuhiro, Y.Nagaoka, S.Takagi and T.Tanzawa** *Prog. Theor. Phys.* **88** (1992) 229–249

[28] **P.H.Morse and H.Feshbach** *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953)

[29] **L.A.Pars** *Quart. J. Mech. and Appl. Math.* **7** (1954) 338–351

[30] **L.A.Pars** *A treatise on analytical dynamics* (Heinemann, London, 1965)

[31] **V.V.Rumiantsev** *PMM* **46** (1982) 1–8

[32] **L.S.Schulman** *Techniques and applications of path integration* (Wiley, New York, 1981)

[33] **K.Sundermeyer** *Constrained dynamics* (Springer-Verlag, Berlin, 1982)

[34] **F.Udwadia and R.Kalaba** *Proc. R. Soc. Lond. A* **439** (1992) 407-410

[35] **E.Zauderer** *Partial differential equations of applied mathematics* (Wiley, Singapore, 1989)

[36] *Encyclopaedia Britannica* (1910) **12** 631