Continuum Modeling and Simulation.

Rodolfo R. Rosales, Adam Powell, Franz-Josef Ulm, Kenneth Beers

MIT, Spring 2006

Contents

Chapter 1

Introduction

1.1 Scope and Purpose of This Document

This document serves as lecture notes for the Continuum Modeling and Simulation set of lectures in Introduction to Modeling and SImulation, Spring, 2006. As with the lectures themselves, this is authored by the six faculty members who teach the twelve lectures in this set. As such, the length, style, and notation may differ slightly from lecturer to lecturer. This document therefore serves not only to provide all of the material in one place, but also to help us to unify the notation schemes as much as possible in order to smooth the transitions as much as possible.

1.2 Lecture Outline

Lectures in this series, and chapters in this document, are summarized as follows:

1. **February 15–17: Rodolfo Rosales.** These lectures will introduce the continuum approximations which roughly describe physical systems at a coarse scale, such as density and displacement fields, and also introduce conservation laws as a framework for continuum modeling. Diffusion and heat conduction will be the motivating examples for this section.

2. **February 21–27: Adam Powell.** The treatment of heat conduction will expand in two ways: first discussing methods for time integration of the transient heat conduction equation, and second introducing the enthalpy method for incorporating phase changes into heat conduction simulations. These lectures will also include a brief introduction to the weighed residual approach to the finite element method.

3. **March 1–March 6: Franz-Josef Ulm.** This section will introduce the variational approach to finite elements for solid mechanics. It will focus on the statics of a gravity dam, illustrating the one-element solution and accuracy as more elements are introduced.

4. **March 8–March 13: Kenneth Beers.** Moving to fluids, this section will discuss the complications inherent to discretizing the Navier-Stokes equations for incompressible fluid flow, including spurious modes in the pressure field and artificial diffusion due to the convective terms, and present methods for resolving these problems.

5. **March 15, 20: Raul Radovitzky.** This section will demonstrate a hybrid Lagrangian and Eulerian approach for modeling fully-coupled interactions between continuum fluids and solids, often referred to as Fluid-Structure Interactions.

6. **March 17: Rodolfo Rosales.** Rosales returns to discuss the relationship between discrete and continuum behavior, focusing on longitudinal and transverse waves and torsion-coupled pendulums.

This document covers the lectures of Rosales, Powell and Ulm, including the final Rosales lecture.

1.3 List of Symbols

Symbols used in more than one section of this document:

Symbol	Name	S.I. Units
T	Temperature	K, °C
c, c_p	Heat capacity[1]	$\frac{\text{J}}{\text{kg}\cdot\text{K}}$
ρ	Density	$\frac{\text{kg}}{\text{m}^3}$
k	Thermal conductivity	$\frac{\text{W}}{\text{m}\cdot\text{K}}$
ν	Thermal diffusivity $= \frac{k}{\rho c}$	$\frac{\text{m}^2}{\text{s}}$
$\mathbf{q}$, q_x	Heat flux vector, x-component	$\frac{\text{W}}{\text{m}^2}$
$\mathbf{u}$	Flow velocity	$\frac{\text{m}}{\text{s}}$
ρ	Density	$\frac{\text{kg}}{\text{m}^3}$
N_i	FEM shape function i	
ξ, η	Local element coordinates	

Chapter 2

Conservation Laws in Continuum Modeling

These notes give examples illustrating how conservation principles are used to obtain (phenomenological) continuum models for physical phenomena. The general principles are presented, with examples from traffic flow, river flows, heat conduction, granular flows, gas dynamics and diffusion.

2.1 Introduction.

In formulating a mathematical model for a continuum physical system, there are three basic steps that are often used:

A. Identify appropriate conservation laws (e.g. mass, momentum, energy, etc) and their corresponding densities and fluxes.

B. Write the corresponding equations using conservation.

C. Close the system of equations by proposing appropriate relationships between the fluxes and the densities.

Of these steps, the mathematical one is the second. While it involves some subtlety, once you understand it, its application is fairly mechanical. The first and third steps involve physical issues, and (generally) the third one is the hardest one, where all the main difficulties appear in developing a new model. In what follows we will go through these steps, using some practical examples to illustrate the ideas.

Of course, once a model is formulated, a **fourth step** arises, which is that of analyzing and validating the model, comparing its predictions with observations ... and correcting it whenever needed. This involves simultaneous mathematical and physical thinking. You should never forget that a model is no better than the approximations (explicit and/or implicit) made when deriving it. It is never a question of just "solving" the equations, forgetting what is behind them.

2.2 Continuum Approximation; Densities and Fluxes.

The modeling of physical variables as if they were a continuum field is almost always an approximation. For example, for a gas one often talks about the density ρ, or the flow velocity $\mathbf{u}$, and thinks of them as functions of space and time: $\rho = \rho(\mathbf{x}, t)$ or $\mathbf{u} = \mathbf{u}(\mathbf{x}, t)$. But the fact is that a gas is made up by very many discrete molecules, and the concepts of density, or flow velocity, only make sense as local averages. These averages must be made over scales large enough that the discreteness of the gas becomes irrelevant, but small enough that the notion of these local averages varying in space and time makes sense.

Thus, **in any continuum modeling there are several scales.** On the one hand one has the **"visible" scales,** which are the ones over which the mathematical variables in the model vary (densities, fluxes). On the other hand, there are the **"invisible" scales,** that pertain to the micro-scales that have been averaged in obtaining the model. **The second set of scales must be much smaller than the first set for the model to be valid.** Unfortunately, this is not always the case, and whenever this fails all sort of very interesting (and largely open) problems in modern science and engineering arise.

Note that the reason people insist on trying to use continuum type models, even in situations where one runs into the difficulties mentioned at the end of the last paragraph, is that continuum models are often much simpler (both mathematically and computationally) than anything else, and supply general understanding that is often very valuable.

The first step in the modeling process is to identify conserved quantities (e.g. mass) and define the appropriate densities and fluxes — as in the following examples.

2.2.1 Examples

Example 2.2.1 River Flow (a one dimensional example).

Consider a nice river (or a channel) flowing down a plain (e.g. the Mississippi, the Nile, etc.). Let x be the length coordinate along the river, and at every point (and time) along the river let $A = A(x, t)$ be the filled (by water) cross-section of the river bed.

We note now that A is the **volume density** *(volume per unit length) of water along the river. We also note that, since water is incompressible,* **volume is conserved**.[1] *Finally, let $Q = Q(x, t)$ be the* **volume flux** *of water down the river (i.e.: volume per unit time). Notice that, if $u = u(x, t)$ is the average* **flow velocity** *down the river, then* $\boxed{Q = uA}$ *(by* definition *of u).*

Thus, in this case, an appropriate conservation law is the **conservation of volume**, *with corresponding density A and flux Q. We note that both A and Q are regularly measured at various points along important rivers.*

Example 2.2.2 Traffic Flow (a one dimensional example).

Consider a one lane road, in a situation where there are no cross-roads (e.g.: a tunnel, such as the Lincoln tunnel in NYC, or the Summer tunnel in Boston). Let x be length along the road. Under "heavy" traffic conditions,[2] *we can introduce the notions of* **traffic density** *$\rho = \rho(x, t)$ (cars per*

[1] We are neglecting here such things as evaporation, seepage into the ground, etc. This cannot always be done.

[2] Why must we assume "heavy" traffic?

unit length) and **traffic flow** $q = q(x,t)$ *(cars per unit time). Again, we have* $\boxed{q = u\rho,}$ *where* u *is the average* **car flow velocity** *down the road.*

In this case, the appropriate conservation law is, obviously, the **conservation of cars.** *Notice that this is one example where the continuum approximation is rather borderline (since, for example, the local averaging distances are almost never much larger than a few car separation lengths). Nevertheless, as we will see, one can gain some very interesting insights from the model we will develop (and some useful practical facts).*

Example 2.2.3 Heat Conduction.

Consider the thermal energy in a chunk of solid material (such as, say, a piece of copper). Then the **thermal energy density** *(thermal energy per unit volume) is given by* $U = c\,T(\mathbf{x},t)$*, where* T *is the temperature,* c *is the specific heat per unit mass, and* ρ *is the density of the material (for simplicity we will assume here that both* c *and* ρ *are constant). The* **thermal energy flow**, $\mathbf{q} = \mathbf{q}(\mathbf{x},t)$ *is now a vector, whose magnitude gives the energy flow across a unit area normal to the flow direction.*

In this case, assuming that heat is not being lost or gained from other energy forms, the relevant conservation law is the **conservation of heat energy.**

Example 2.2.4 Steady State (dry) Granular Flow.

Consider steady state (dry) granular flow down some container (e.g. a silo, containing some dry granular material, with a hole at the bottom). At every point we characterize the flow in terms of two velocities: an **horizontal (vector) velocity** $\mathbf{u} = \mathbf{u}(x,y,z,t)$*, and a* **vertical (scalar) velocity** $v = v(x,y,z,t)$*, where* x *and* y *are the horizontal length coordinates, and* z *is the vertical one.*

The **mass flow rate** *is then given by* $\mathbf{Q} = \rho\,[\mathbf{u}, v]$*, where* ρ *is the* **mass density** *— which we will assume is nearly constant. The relevant conservation is now the* **conservation of mass.**

This example is different from the others in that we are looking at a steady state situation. We also note that this is another example where the continuum approximation is quite often "borderline", since the scale separation between the grain scales and the flow scales is not that great.

Example 2.2.5 Inviscid Fluid Flow.

For a fluid flowing in some region of space, we consider now two conservation laws: **conservation of mass** *and* **conservation of linear momentum***. Let now* $\rho = \rho(\mathbf{x},t)$, $\mathbf{u} = \mathbf{u}(\mathbf{x},t)$ *and* $p = p(\mathbf{x},t)$ *be, respectively, the fluid density, flow velocity, and pressure — where we use either* $[u,v,w]$ *or* $[u_1,u_2,u_3]$ *to denote the components of* $\mathbf{u}$*, and either* $[x,y,z]$ *or* $[x_1,x_2,x_3]$ *to denote the components of* $\mathbf{x}$*. Then:*

- *The* **mass conservation law density** *is* .... ρ.
- *The* **mass conservation law flow** *is* .... $\rho\,\mathbf{u}$.
- *The* **linear momentum conservation law density** *is* .... $\rho\,\mathbf{u}$.
- *The* **linear momentum conservation law flow** *is* .... $\rho\,\mathbf{u}\otimes\mathbf{u} + p\,\mathbf{I}$.

The first two expressions above are fairly obvious, but the last two (in particular, the last one) require some explanation. First of all, momentum is a vector quantity. Thus its conservation is

equivalent to three conservation laws, with a vector density and a rank two tensor[3] *flow (we explain this below). Second, momentum can be transferred from one part of a liquid to another in two ways:*
Advection: *as a parcel of fluid moves, it carries with it some momentum. Let us consider this mechanism component by component: The momentum density component* $\rho\, u_i$ *is advected with a flow rate* $\rho\, u_i\, \mathbf{u} = \rho\, [u_i u_1, u_i u_2, u_i u_3]$. *Putting all three components together, we get for the momentum flux (due to advection) the expression* $\rho\, [u_i\, u_j] = \rho\, \mathbf{u} \otimes \mathbf{u}$ *— i.e., a rank two tensor, where each row (freeze the first index) corresponds to the flux for one of the momentum components.*
Forces: *momentum is transferred by the forces exerted by one parcel of fluid on another. If we assume that the fluid is inviscid, then these forces can only be normal, and are given by the pressure (this is, actually, the "definition" of inviscid). Thus, again, let us consider this mechanism component by component: the momentum transfer by the pressure in the direction given by the unit vector*[4] $\mathbf{e}_i = [\delta_{i\,j}]$, *corresponding to the density* $\rho\, u_i$, *is the force per unit area (normal to* $\mathbf{e_i}$*) by the fluid. Thus the corresponding momentum flow vector is* $p\, \mathbf{e}_i$. *Putting all three components together, we get for the momentum flux (due to pressure forces) the expression* $p\, [\delta_{i\,j}] = p\, \mathbf{I}$ *— again a rank two tensor, now a scalar multiple of the identity rank two tensor* $\mathbf{I}$.

Regarding the zero viscosity (inviscid) assumption: Fluids can also exert tangential forces, which also affect the momentum transfer. Momentum can also be transferred in the normal direction by diffusion of "faster" molecules into a region with "slower" molecules, and vice versa. Both these effects are characterized by the viscosity coefficient — which here we assume can be neglected.

Note that in some of the examples we have given only one conservation law, and in others two (further examples, with three or more conservation laws invoked, exist). The reason will become clear when we go to the third step (step **C** in section 2.1). In fact, steps **A** and **C** in section 2.1 are intimately linked, as we will soon see.

2.3 Conservation Laws in Mathematical Form.

In this section we assume that we have identified some conservation law, with conserved density $\rho = \rho(\mathbf{x}, t)$, and flux $\mathbf{F} = \mathbf{F}(\mathbf{x}, t)$, and derive mathematical formulations for the conservation hypothesis. In other words, we will just state in mathematical terms the fact that ρ is the density for a conserved quantity, with flux $\mathbf{F}$.

First consider the one dimensional case (where the flux F is a scalar, and there is only one space coordinate: x). In this case, consider some (fixed) arbitrary interval in the line $\Omega = \{a \le x \le b\}$, and let us look at the evolution in time of the conserved quantity inside this interval. At any given time, the total amount of conserved stuff in Ω is given by (this by definition of density)

$$M(t) = \int_a^b \rho(x, t)\, dx\,. \tag{3.1}$$

Further, the net rate at which the conserved quantity enters Ω is given by (definition of flux)

$$R(t) = F(a, t) - F(b, t)\,. \tag{3.2}$$

[3] If you do not know what a tensor is, just think of it as a vector with more than one index (the rank is the number of indexes). This is all you need to know to understand what follows.

[4] Here $\delta_{i\,j}$ is the Kronecker delta, equal to 1 if $i = j$, and to 0 if $i \neq j$.

It is also possible to have **sources and sinks** for the conserved quantity.[5] In this case let $\boxed{s = s(x,t)}$ be the total net amount of the conserved quantity, per unit time and unit length, provided by the sources and sinks. For the interval Ω we have then a net rate of added conserved stuff, per unit time, given by

$$S(t) = \int_a^b s(x,t)\,dx\,. \tag{3.3}$$

The conservation law can now be stated in the mathematical form

$$\frac{d}{dt}M = R + S\,, \tag{3.4}$$

which **must apply for any choice of interval Ω.** Since this equation involves only integrals of the relevant densities and fluxes, it is known as the **Integral Form of the Conservation Law.**

Assume now that the densities and fluxes are nice enough to have nice derivatives. Then we can write:

$$\frac{d}{dt}M = \int_a^b \frac{\partial}{\partial t}\rho(x,t)\,dx \quad \text{and} \quad R = -\int_a^b \frac{\partial}{\partial x}F(x,t)\,dx\,. \tag{3.5}$$

Equation (3.4) can then be re-written in the form

$$\int_a^b \left(\frac{\partial}{\partial t}\rho(x,t) + \frac{\partial}{\partial x}F(x,t) - s(x,t)\right)\,dx = 0\,, \tag{3.6}$$

which must apply for any choice of the interval Ω. It follows that the integrand above in (3.6) must vanish identically. This then yields the following partial differential equation involving the density, flux and source terms:

$$\frac{\partial}{\partial t}\rho(x,t) + \frac{\partial}{\partial x}F(x,t) = s(x,t)\,. \tag{3.7}$$

This equation is known as the **Differential Form of the Conservation Law.**

Remark 2.3.1 *You may wonder why we even bother to give a name to the form of the equations in (3.4), since the differential form in (3.7) appears so much more convenient to deal with (it is just one equation, not an equation for every possible choice of Ω). The reason is that it is not always possible to assume that the densities and fluxes have nice derivatives. Oftentimes the physical systems involved develop, as they evolve,[6] short enough scales that force the introduction of discontinuities into the densities and fluxes — and then (3.7) no longer applies, but (3.4) still does.* **Shock waves** *are the best known example of this situation. Examples of shock waves you may be familiar with are: the sonic boom produced by a supersonic aircraft; the hydraulic jump occurring near the bottom of the discharge ramp in a large dam; the wave-front associated with a flood moving down a river; the backward facing front of a traffic jam; etc. Some shock waves can cause quite spectacular effects, such as those produced by supernova explosions.*

[5] As an illustration, in the inviscid fluid flow case of example 2.2.5, the effects of gravity translate into a vertical source of momentum, of strength $\rho\, g$ per unit volume — where g is the acceleration of gravity. Other body forces have similar effects.

[6] Even when starting with very nice initial conditions.

Now let us consider the multi-dimensional case, when the flux $\mathbf{F}$ is a vector. In this case, consider some (fixed but arbitrary) region in space Ω, with boundary $\partial\Omega$, and inside unit normal along the boundary $\hat{\mathbf{n}}$. We will now look at the evolution in time of the conserved quantity inside this region. At any given time, the total amount of conserved stuff in Ω is given by

$$M(t) = \int_{\Omega} \rho(\mathbf{x}, t)\, dV\,. \tag{3.8}$$

On the other hand, the net rate at which the conserved quantity enters Ω is given by

$$R(t) = \int_{\partial\Omega} \mathbf{F}(\mathbf{x}, t) \cdot \hat{\mathbf{n}}\, dS\,. \tag{3.9}$$

Let also $\boxed{s = s(\mathbf{x}, t)}$ be the total net amount of conserved quantity, per unit time and unit volume, provided by any sources and/or sinks. For the region Ω we have then a net rate of added conserved stuff, per unit time, given by

$$S(t) = \int_{\Omega} s(\mathbf{x}, t)\, dV\,. \tag{3.10}$$

The conservation law can now be stated in the mathematical form (compare with equation (3.4)) — **Integral Form of the Conservation Law:**

$$\frac{d}{dt} M = R + S\,, \tag{3.11}$$

which **must apply for any choice of the region Ω.**

If the densities and fluxes are nice enough to have nice derivatives, we can write:

$$\frac{d}{dt} M = \int_{\Omega} \frac{\partial}{\partial t} \rho(\mathbf{x}, t)\, dV \quad \text{and} \quad R = -\int_{\Omega} \mathsf{div}(\mathbf{F}(\mathbf{x}, t))\, dV\,, \tag{3.12}$$

where we have used the Gauss divergence theorem for the second integral. Equation (3.11) can then be re-written in the form

$$\int_{\Omega} \left(\frac{\partial}{\partial t} \rho(\mathbf{x}, t) + \mathsf{div}(\mathbf{F}(\mathbf{x}, t)) - s(\mathbf{x}, t) \right) dV = 0\,, \tag{3.13}$$

which must apply for any choice of the region Ω. It follows that the integrand above in (3.13) must vanish identically. This then yields the following partial differential equation involving the density, flux and source terms (compare with equation (3.7))

$$\frac{\partial}{\partial t} \rho(\mathbf{x}, t) + \mathsf{div}(\mathbf{F}(\mathbf{x}, t)) = s(\mathbf{x}, t)\,. \tag{3.14}$$

This equation is known as the **Differential Form of the Conservation Law.**

Remark 2.3.2 *In the case of a* **vector conservation law***, the density ρ and the source term s will both be vectors, while the flux $\mathbf{F}$ will be a rank two tensor (each row being the flux for the corresponding element in the density vector ρ). In this case equation (3.14) is valid component by*

component, but can be given a vector meaning if we define the divergence for a rank two tensor $\mathbf{F} = [F_{ij}]$ *as follows:*

$$\mathsf{div}(\mathbf{F}) = \left[\sum_j \frac{\partial}{\partial x_j} F_{ij} \right] ,$$

so that $\mathsf{div}(\mathbf{F})$ *is a vector (each element corresponding to a row in* $\mathbf{F}$*). You should check that this is correct.*[7]

2.4 Phenomenological Equation Closure.

From the results in section 2.3 it is clear that each conservation principle can be used to yield an evolution equation relating the corresponding density and flux. However, this is not enough to provide a complete system of equations, since each conservation law provides only one equation, but requires two (in principle) "independent" variables. Thus extra relations between the fluxes and the densities must be found to be able to formulate a complete mathematical model. This is the **Closure Problem**, and it often requires making further assumptions and approximations about the physical processes involved.

Closure is actually the hardest and the subtler part of any model formulation. How good a model is, typically depends on how well one can do this part. Oftentimes the physical processes considered are very complex, and no good understanding of them exist. In these cases one is often forced to make "brute force" phenomenological approximations (some formula — with a few free parameters — relating the fluxes to the densities is proposed, and then it is fitted to direct measurements). Sometimes this works reasonably well, but just as often it does not (producing situations with very many different empirical fits, each working under some situations and not at all in others, with no clear way of knowing "a priori" if a particular fit will work for any given case).

We will illustrate how one goes about resolving the closure problem using the examples introduced earlier in subsection 2.2.1. These examples are all "simple", in the sense that one can get away with algebraic formulas relating the fluxes with the densities. However, this is not the only possibility, and situations where extra differential equations must be introduced also arise. The more complex the process being modeled is, the worse the problem, and the harder it is to close the system (with very many challenging problems still not satisfactorily resolved).

An important point to be made is that **the formulation of an adequate mathematical model is only the beginning.** As the examples below will illustrate, it is often the case that the mathematical models obtained are quite complicated (reflecting the fact that the phenomena being modeled are complex), and often poorly understood. Thus, even in cases where accurate mathematical models have been known for well over a century (as in classical fluids), there are plenty of open problems still around ... and even now new, un-expected, behaviors are being discovered in experimental laboratories. The fact is that, for these complex phenomena, mathematics alone is not enough. There is just too much that can happen, and the equations are too complicated to have explicit solutions. The only possibility of advance is by a simultaneous approach incorporating experiments and observations, numerical calculations, and theory.

[7] Recall that, for a vector field, $\mathsf{div}(\mathbf{v}) = \sum_j \frac{\partial}{\partial x_j} v_j$.

2.4.1 Examples

Example 2.4.1 River Flow (see example 2.2.1).

In this case we can write the conservation equation

$$A_t + Q_x = 0\,, \tag{4.1}$$

where A and Q were introduced in example 2.2.1, and we ignore any sources or sinks for the water in the river. In order to close the model, we now claim that it is reasonable to assume that Q is a function of A; that is to say $\boxed{Q = Q(A,x)}$ *— for a uniform, man-made channel, one has* $\boxed{Q = Q(A).}$ *We justify this hypothesis as follows:*

First: *For a given river bed shape, when the flow is steady (i.e.: no changes in time) the average flow velocity u follows from the balance between the force of gravity pulling the water down the slope, and the friction force on the river bed. This balance depends only on the river bed shape, its slope, and how much water there is (i.e. A). Thus, under these conditions, we have $u = u(A,x)$. Consequently $Q = Q(A,x) = u(A,x)\,A$.*

Second: *As long as the flow in the river does not deviate too much from steady state ("slow" changes), the we can assume that the relationship $Q = Q(A,x)$ that applies for steady flow remains (approximately) valid. This is the* **quasi-equilibrium approximation***, which is often invoked in problems like this. How well it works in any given situation depends on how fast the processes leading to the equilibrium situation (the one that leads to $Q = Q(A,x)$) work — relative to the time scales of the river flow variations one is interested in. For actual rivers and channels, it turns out that this approximation is good enough for many applications.*

Of course, the actual functional relationship $Q = Q(A,x)$ (to be used to model a specific river) cannot be calculated theoretically, and must be extracted from actual measurements of the river flow under various conditions. The data is then fitted by (relatively simple) empirical formulas, with free parameters selected for the best possible match.

However, it is **possible to get a qualitative idea of roughly how Q depends on A,** *by the following simple argument: The force pulling the water downstream (gravity) is proportional to the slope of the bed, the acceleration of gravity, the density of water, and the volume of water. Thus, roughly speaking, this force has the form $F_g \approx c_g\,A$ (where $c_g = c_g(x)$ is some function). On the other hand, the force opposing this motion, in the simplest possible model, can be thought as being proportional to the wetted perimeter of the river bed (roughly $P \propto \sqrt{A}$) times the frictional force on the bed (roughly proportional to the velocity u). That is $F_f \approx c_f\,u\sqrt{A}$, for some friction coefficient c_f. These two forces must balance ($F_g = F_f$), leading to $u \approx c_u\sqrt{A}$ (where $c_u = c_g/c_f$), thus:*

$$Q \approx c_u\,A^{3/2}\,. \tag{4.2}$$

Of course, this is too simple for a real river. But the feature of the flux increasing faster than linear is generally true — so that Q **as a function of A produces a concave graph,** *with* $\boxed{dQ/dA > 0}$ *and* $\boxed{d^2Q/dA^2 > 0.}$

Example 2.4.2 Traffic Flow (see example 2.2.2).

In this case we can write the conservation equation

$$\rho_t + q_x = 0\,, \tag{4.3}$$

where ρ and q were introduced in example 2.2.2, and we ignore any sources or sinks for cars (from road exit and incoming ramps, say). Just as in the river model, we close now the equations by claiming that it is reasonable to assume that q is a function of ρ, that is to say $\boxed{q = q(\rho, x)}$ *— for a nice, uniform, road, one has* $\boxed{q = q(\rho).}$ *Again, we use a* **quasi-equilibrium approximation** *to justify this hypothesis:*

Under steady traffic conditions, it is reasonable to assume that the drivers will adjust their car speed to the local density (drive faster if there are few cars, slower if there are many). This yields $u = u(\rho, x)$, thus $q = u(\rho, x)\rho = q(\rho, x)$. Then, if the traffic conditions do not vary too rapidly, we can assume that the equilibrium relationship $q = q(\rho, x)$ will still be (approximately) valid — quasi-equilibrium approximation.

As in the river flow case, the actual functional dependence to be used for a given road must follow from empirical data. Such a fit for the Lincoln tunnel in NYC is given by[8]

$$q = a\,\rho\,\log(\rho_j/\rho)\,, \tag{4.4}$$

where $a = 17.2$ mph, and $\rho_j = 228$ vpm (vehicles per mile). The generic shape of this formula is always true: q is a convex function of ρ, reaching a maximum flow rate q_m for some value $\rho = \rho_m$, and then decreases back to zero flow at a jamming density $\rho = \rho_j$. In particular, $dq/d\rho$ **is a decreasing function of ρ, with** $\boxed{d^2q/d\rho^2 < 0.}$

For the formula above in (4.4), we have: $\rho_m = 83$ vpm and $q_m = 1430$ vph (vehicles per hour), with a corresponding flow speed $u_m = q_m/\rho_m = a$. The very existence of ρ_m teaches us a **rather useful fact**, *even before we solve any equation: in order to maximize the flow in a highway, we should try to keep the car density near the optimal value ρ_m. This is what the lights at the entrances to freeways attempt to do during rush hour. Unfortunately, they do not work very well for this purpose, as some analysis with the model above (or just plain observation of an actual freeway) will show.*
In this example the continuum approximation is rather borderline. Nevertheless, the equations have the right qualitative (and even rough quantitative) behavior, and are rather useful to understand many features of how heavy traffic behaves.

Example 2.4.3 Heat Conductivity (see example 2.2.3).

In this case we can write the conservation equation

$$\rho\,c\,T_t + \mathsf{div}(\mathbf{q}) = s\,, \tag{4.5}$$

where $\rho\,c$, T and $\mathbf{q}$ were introduced in example 2.2.3, and $s = s(\mathbf{x}, t)$ is the heat supplied (per unit volume and unit time) by any sources (or sinks) — e.g. electrical currents, chemical reactions, etc.

[8] Greenberg, H., 1959. An analysis of traffic flow. *Oper. Res.* **7:**79–85.

We now complete the model by observing that heat flows from hot to cold, and postulating that the heat flow across a temperature jump is proportional to the temperature difference (this can be checked experimentally, and happens to be an accurate approximation). This leads to **Fick's Law** *for the heat flow:*

$$\mathbf{q} = -k\,\nabla T\,, \tag{4.6}$$

where k is the **coefficient of thermal conductivity** *of the material.*[9] *For simplicity we will assume here that c, ρ and k are constant — though this is not necessarily true in general.*

Substituting (4.6) into (4.5), we then obtain the **heat or diffusion equation:**

$$T_t = \nu\,\nabla^2 T + f\,, \tag{4.7}$$

where $\nu = \dfrac{k}{\rho c}$ is the **thermal diffusivity** *of the material, and $f = \dfrac{s}{\rho c}$.*

In deriving the equation above, we assumed that the heat was contained in a chunk of solid material. The reason for this is that, in a fluid, heat can also be transported by motion of the fluid (convection). In this case (4.6) above must be modified to:

$$\mathbf{q} = -k\,\nabla T + \rho\, c\, T\,\mathbf{u}\,, \tag{4.8}$$

where $\mathbf{u} = \mathbf{u}(\mathbf{x}, t)$ is the fluid velocity. Then, instead of (4.7), we obtain

$$T_t + \mathsf{div}(\mathbf{u}T) = \nu\,\nabla^2 T + f\,. \tag{4.9}$$

In fact, this is the simplest possible situation that can occur in a fluid. The reason is that, generally, the fluid density depends on temperature, so that the fluid motion ends up coupled to the temperature variations, due to buoyancy forces. Then equation (4.9) must be augmented with the fluid equations, to determine $\mathbf{u}$ and the other relevant fluid variables — see example 2.4.5.

Remark 2.4.1 *Note that ν has dimensions $\dfrac{Length^2}{Time}$. Thus, given a length L, a time scale is provided by $\tau = L^2/\nu$. Roughly speaking, this is the amount of time it would take to heat (or cool) a region of size L by diffusion alone. If you go and check the value of ν for (say) water, you will find out that it would take a rather long time to heat even a cup of tea by diffusion alone* (you should do this calculation). *The other term in (4.9) is crucial in speeding things up.*

Remark 2.4.2 *If the fluid is incompressible, then $\mathsf{div}(\mathbf{u}) = 0$ (see example 2.4.5), and equation (4.9) takes the form*

$$T_t + (\mathbf{u}\cdot\nabla)T = \nu\,\nabla^2 T + f\,. \tag{4.10}$$

Note that the left hand side in this equation is just the time derivative of the temperature in a fixed parcel of fluid, as it is being carried around by the flow.

Remark 2.4.3 *Equations such as (4.9) and (4.10) are satisfied not just by the temperature, but by many other quantities that propagate by diffusion (i.e.: their fluxes satisfy Fick's Law (4.6)). Examples are given by any chemicals in solution in a liquid (salt, sugar, colorants, pollutants, etc.). Of course, if there are any reactions these chemicals participate in, these reactions will have to be incorporated into the equations (as sources and sinks).*

[9] k must be measured experimentally, and varies from material to material.

Example 2.4.4 Steady State (dry) Granular Flow (see example 2.2.4).

In this case we can write the conservation equation

$$\mathsf{div}(\mathbf{Q}) = 0\,, \tag{4.11}$$

where $\mathbf{Q} = \rho[\mathbf{u}, v]$ *is as in example 2.2.4, and there are no time derivatives involved because we assumed that the density* ρ *was nearly constant (we also assume that there are no sources or sinks for the media). These equation involves three unknowns (the three flow velocities), so we need some extra relations between them to close the equation.*

The argument now is as follows: as the grain particles flow down (because of the force of gravity), they will also — more or less randomly — move to the sides (due to particle collisions). We claim now that, on the average, it is easier for a particle to move from a region of low vertical velocity to one of high vertical velocity than the reverse.[10] *The simplest way to model this idea is to propose that the horizontal flow velocity* $\mathbf{u}$ *is proportional to the horizontal gradient of the vertical flow velocity* v*. Thus we propose a law of the form:*

$$\mathbf{u} = b\,\nabla_\perp v \tag{4.12}$$

where b *is a* **coefficient (having length dimensions)** *and* $\nabla_\perp$ *denotes the* **gradient with respect to the horizontal coordinates** x **and** y**.** *Two important points:*

- **A.** *Set the coordinate system so that the* z *axis points down. Thus* v *is positive when the flow is downwards, and* b **above is positive**.
- **B.** *Equation (4.12) is a purely empirical proposal, based on some rough intuition and experimental observations. However, it works. The predictions of the resulting model in equation (4.13) below have been checked against laboratory experiments, and they match the observations, provided that the value of* b *is adjusted properly (typically,* b *must be taken around a few particle diameters).*

Substituting (4.12) into (4.11), using the formula for the divergence, and eliminating the common constant factor ρ*, we obtain the following model equation for the vertical velocity* v*:*

$$0 = v_z + b\,\nabla_\perp^2 v = v_z + b\,(v_{xx} + v_{yy})\,. \tag{4.13}$$

Note that this is a diffusion equation, except that the role of time has been taken over by the vertical coordinate z*. Mathematical analysis of this equation shows that* **it only makes sense to solve it for** z **decreasing; i.e.: from bottom to top in the container where the flow takes place**. *This, actually, makes perfect physical sense: if you have a container full of (say) dry sand, and you open a hole at the bottom, the motion will propagate upwards through the media. On the other hand, if you move the grains at the top, the ones at the bottom will remain undisturbed. In other words, information about motion in the media propagates upward, not downwards.*

[10] Intuitively: where the flow speed is higher, there is more space between particles where a new particle can move into.

Example 2.4.5 Inviscid Fluid Flow (see example 2.2.5).

In this case, using the densities and fluxes introduced in example 2.2.5, we can write the conservation equations:

$$\rho_t + \mathsf{div}(\rho\,\mathbf{u}) = 0 \tag{4.14}$$

for the **conservation of mass**, *and*

$$(\rho\,\mathbf{u})_t + \mathsf{div}(\rho\,\mathbf{u}\otimes\mathbf{u}) + \nabla p = \mathbf{F} \tag{4.15}$$

for the **conservation of momentum.** *Here* $\mathbf{F} = \mathbf{F}(\mathbf{x}, t)$ *denotes the body forces*[11] *(which are momentum sources), and we have used the mathematical identity* (you should check this) $\mathsf{div}(p\,\mathbf{I}) = \nabla p$. *Another easy to check mathematical identity is* $\mathsf{div}(\mathbf{u}\otimes\mathbf{m}) = (\mathsf{div}(\mathbf{m}))\,\mathbf{u} + (\mathbf{m}\cdot\nabla)\,\mathbf{u}$. *Using this second identity, with* $\mathbf{m} = \rho\,\mathbf{u}$, *in equation (4.15), and substituting from equation (4.14) to eliminate the term containing the divergence of* $\mathbf{m}$, *we obtain:*

$$\rho\,(\mathbf{u}_t + (\nabla\cdot\mathbf{u})\,\mathbf{u}) + \nabla p = \mathbf{F}\,. \tag{4.16}$$

The problem now is that we have four equations and five unknowns (density, pressure and the three velocities). **An extra equation is needed.** *Various possibilities exist, and we illustrate a few below.*

Incompressibility Assumption (liquids).

Liquids are generally very had to compress. This means that, as a parcel of fluid is carried around by the flow, its volume (equivalently, its density) will change very little. If we then make the assumption that the liquid density does not change at all (due to pressure changes ... it certainly may change due to temperature changes, or solutes[12] *in the liquid), then we obtain the following additional equation:*

$$\rho_t + (\nabla\cdot\mathbf{u})\,\rho = 0\,. \tag{4.17}$$

This equation simply states that the time derivative of the density, following a parcel of fluid as it moves, vanishes. In other words: the fluid is incompressible (though it need not have a constant density). In this case we can write a complete system of equations for the fluid motion. Namely:

$$\left.\begin{array}{rcl} 0 &=& \rho_t + (\nabla\cdot\mathbf{u})\,\rho\,, \\ 0 &=& \mathsf{div}(\mathbf{u})\,, \\ \mathbf{F} &=& \rho\,(\mathbf{u}_t + (\nabla\cdot\mathbf{u})\,\mathbf{u}) + \nabla p\,, \end{array}\right\} \tag{4.18}$$

where the second equation follows from (4.14), upon use of (4.17). These are known as the **Incompressible Euler Equations** *for a fluid. The "simplest" situation arises when* ρ *can be assumed constant, and then the first equation above is not needed. However, even in this case, the behavior of the solutions to these equations is not well understood — and extremely rich.*

Remark 2.4.4 *The equations above ignore viscous effects, important in modeling many physical situations. Viscosity is incorporated with the method used in example 2.4.3, by adding to the momentum flux components proportional to derivatives of the flow velocity* $\mathbf{u}$. *What results from this are the* **Incompressible Navier-Stokes Equations.**

Furthermore, heat conduction effects can also be considered (and are needed to correctly model many physical situations). This requires the introduction of a new independent variable into the equations (temperature), and the use of one more conservation law (energy).

[11] Such as gravity.

[12] For example, salt.

Gas Dynamics.

For gases one cannot assume incompressibility. In this case, one must introduce another conservation law **(conservation of energy),** *and yet another variable: the* **internal energy per unit mass** e**.** *This results in five equations (conservation of mass (4.14), conservation of momentum (4.15), and conservation of energy) and six variables (density ρ, flow velocity* $\mathbf{u}$*, pressure p and internal energy e). At this stage* **thermodynamics** *comes to the rescue, providing an extra relationship:* **the equation of state.** *For example, for an* **ideal gas with constant specific heats (polytropic gas)** *one has:*

$$e = c_v\, T \quad \textit{and} \quad p = R\,\rho\, T \quad \Longrightarrow \quad \textbf{Equation of state:} \quad \boxed{e = \frac{p}{(\gamma - 1)\,\rho},} \tag{4.19}$$

where c_v is the **specific heat at constant volume**, *c_p is the* **specific heat at constant pressure**, *$R = c_p - c_v$ is the* **gas constant** *and $\gamma = c_p/c_v$ is the ratio of specific heats.*

A simplifying assumption *that can be made, applicable in some cases, is that the* **flow is isentropic.**[13] *In this case the pressure is a function of the density only, and (4.14) and (4.15) then form a complete system: the* **Isentropic Euler Equations of Gas Dynamics.** *For a* **polytropic gas:**

$$p = \kappa\, \rho^{\gamma}\,, \tag{4.20}$$

where κ is a constant. In one dimension the equations are

$$\rho_t + (\rho\, u)_x = 0 \quad \textit{and} \quad (\rho\, u)_t + (\rho\, u^2 + p)_x = 0\,, \tag{4.21}$$

where $p = p(\rho)$.

Remark 2.4.5 *The closure problem in this last example involving gas dynamics seemed rather simple, and (apparently) we did not have to call upon any "quasi-equilibrium" approximation, or similar. However, this is so only because we invoked an already existing (mayor) theory: thermodynamics. In effect, in this case, one cannot get closure unless thermodynamics is developed first (no small feat). Furthermore: in fact,* a quasi-equilibrium approximation is involved. Formulas such as the ones above in (4.19, apply only for equilibrium thermodynamics! *Thus, the closure problem for this example is resolved in a fashion that is exactly analogous to the one used in several of the previous examples.*

Remark 2.4.6 *In the fashion similar to the one explained in remark 2.4.4 for the incompressible case, viscous and heat conduction effects can be incorporated into the equations of Gas Dynamics. The result is the* **Navier-Stokes Equations for Gas Dynamics.**

2.5 Concluding Remarks.

Here we have presented the derivation (using conservation principles) of a few systems of equations used in the modeling of physical phenomena. The study of these equations, and of the physical phenomena they model, on the other hand, would require several lifetimes (and is still proceeding). In particular, notice that here we have not even mentioned the **very important subject of boundary conditions** (what to do at the boundaries of, say, a fluid). This introduces a whole set of new complications, and physical effects (such as surface tension).

[13] That is: the entropy is the same everywhere.

Chapter 3

Timestepping Algorithms and the Enthalpy Method

This chapter will expand on the previous description of finite difference discretization of time in the heat equation by providing a framework for implicit and explicit time stepping algorithms. It will then discuss the enthalpy method for tracking phase boundaries e.g. during melting and solidification.

3.1 Introduction

Rodolfo Rosales began this series by discussing conservation laws, with various examples including solute and thermal diffusion. The approach was to treat various phenomena in terms of fields in space and time: a concentration field, a temperature field, etc., with partial differential equations describing the changes in those fields over time.
This section introduces a numerical method called Finite Differences for approximately solving the partial differential equations to give an estimate for the fields themselves. The heat equation is used as an example, with multiple different schemes discussed for time integration (of which explicit time stepping is the only one you will be required to know). This simple but limited approach serves to introduce several aspects of numerical solution of PDEs, from stability of integration methods, to linearization of the equation system into a matrix equation; these will inform your understanding of more powerful but complex methods later.

3.2 Finite Differences and the Energy Equation

The laws of thermodynamics give rise to a general equation for heat conduction given in section 2.4.3 equation 4.7:

$$\rho c \frac{\partial T}{\partial t} = k \frac{\partial^2 T}{\partial x^2} + s. \tag{2.1}$$

There are many analytical solutions to this equation for various initial and boundary conditions; if generation is zero then Fourier series provide a solution for any initial condition; Green's function integrals solve the steady-state equation straightforwardly for any generation expression (and the time-dependent equation somewhat less straightforwardly). But let's face it, these things are a pain,

it's a lot easier to throw the equations at a computer and let the machine do the work. Furthermore, for complex boundary conditions and/or geometries, one is not guaranteed to obtain such a solution at all; Green's functions often don't integrate well, either analytically or numerically.
Into this need steps the simple computational method of finite differences. To use finite differences, we start by discretizing space and time into a finite number of points, as illustrated in figure 2.1. In one dimension, we can call the spatial points x_i, and the time values t_n. For simplicity, here we will consider only uniform discretization, where we define $\Delta x = x_1 - x_0 = x_{i+1} - x_i$ for all i, and $\Delta t = t_{n+1} - t_n$ for all n. The temperature at position x_i and time t_n can be written $T_{i,n}$.

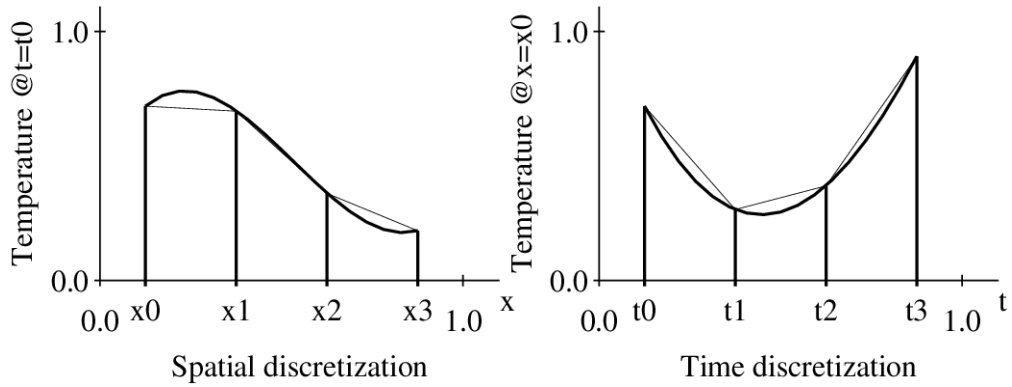

Figure 2.1: Finite difference discretization of space and time. The thin line segments represent the finite difference approximation of the function.

We can then estimate the temperature derivatives as follows:

$$\left.\frac{\partial T}{\partial t}\right|_{x_i,t_{n+1/2}} \simeq \frac{T_{i,n+1} - T_{i,n}}{\Delta t}, \tag{2.2}$$

$$\left.\frac{\partial T}{\partial x}\right|_{x_{i+1/2},t_n} \simeq \frac{T_{i+1} - T_i}{\Delta x} \tag{2.3}$$

The second derivative $\partial^2 T/\partial x^2$ can then be approximated as a derivative of derivatives:

$$\left.\frac{\partial^2 T}{\partial x^2}\right|_{x_i,t_n} \simeq \frac{\left.\frac{\partial T}{\partial x}\right|_{x_{i+1/2},t_n} - \left.\frac{\partial T}{\partial x}\right|_{x_{i-1/2},t_n}}{\Delta x} \simeq \frac{T_{i-1,n} - 2T_{i,n} + T_{i+1,n}}{\Delta x^2} \tag{2.4}$$

3.2.1 Explicit time stepping

These derivative approximations can then be used to estimate solutions to the heat equation (equation 2.1), turning that differential equation into a *difference equation*:

$$\rho c \frac{T_{i,n+1} - T_{i,n}}{\Delta t} = k \frac{T_{i-1,n} - 2T_{i,n} + T_{i+1,n}}{\Delta x^2} + s. \tag{2.5}$$

We can simplify this slightly by recalling the definitions of the thermal diffusivity $\nu = k/\rho c$ and normalized source $f = s/\rho c$ (section 2.4.3, page 15), and defining the *mesh Fourier number* Fo_M as:

$$\mathrm{Fo}_M = \frac{k\Delta t}{\rho c \Delta x^2} = \frac{\nu \Delta t}{\Delta x^2} \tag{2.6}$$

(recall ν is the thermal diffusivity $k/\rho c$), so a rearranged equation 2.5 becomes:

$$T_{i,n+1} = T_{i,n} + \Delta t \left[\nu \frac{T_{i-1,n} - 2T_{i,n} + T_{i+1,n}}{\Delta x^2} + \frac{s}{\rho c} \right] = T_{i,n} + \mathrm{Fo}_M (T_{i-1,n} - 2T_{i,n} + T_{i+1,n}) + f\Delta t \quad (2.7)$$

This gives us a nice algorithm for computing the temperatures at the next timestep from those of the previous timestep and those on the boundaries. This algorithm is called *explicit timestepping*, or the *Forward Euler* timestepping algorithm. It is so straightforward that we can even do it in a simple spreadsheet. But its key drawback is that for large time step sizes, it is unstable, as discussed below.

Explicit timestepping stability criterion

We can regroup the terms on the right side of equation 2.7 to express the temperature in the new timestep as follows:

$$T_{i,n+1} = T_{i,n}(1 - 2\mathrm{Fo}_M) + 2\mathrm{Fo}_M \frac{T_{i-1,n} + T_{i+1,n}}{2} + f\Delta t. \quad (2.8)$$

Neglecting the last term for generation, this is effectively a weighted average between $T_{i,n}$ and the mean of its two neighbors: if $\mathrm{Fo}_M = 0$, then $T_{i,n+1} = T_{i,n}$ (goes nowhere); if $\mathrm{Fo}_M = \frac{1}{2}$, then $T_{i,n+1} = \frac{1}{2}(T_{i-1,n} + T_{i+1,n})$ (mean of the neighbors); if Fo_M is between 0 and $\frac{1}{2}$ then $T_{i,n+1}$ will be between these two. But if $\mathrm{Fo}_M > \frac{1}{2}$, then the new temperature goes beyond the mean of the neighbors. This makes the solution unstable, as oscillations in the solution will grow geometrically with each timestep, as shown in figure 2.2.

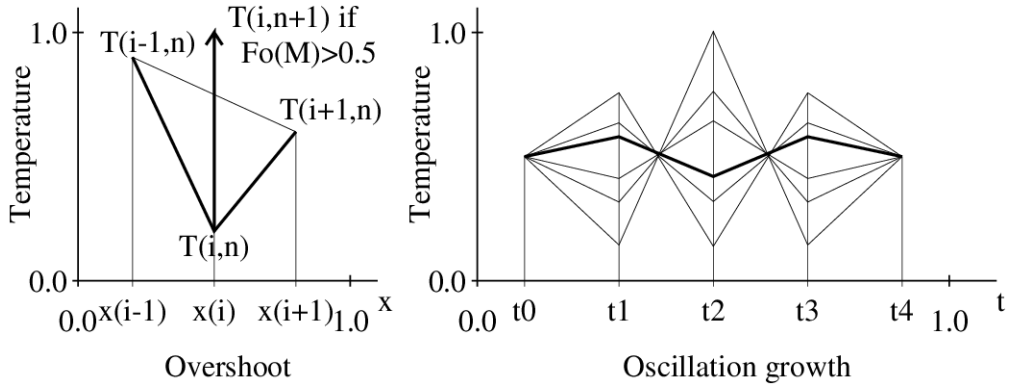

Figure 2.2: "Overshoot" of the average neighboring temperature for $\mathrm{Fo}_M > 0.5$; growth of an oscillation for $\mathrm{Fo}_M = 0.7$ with T_0 and T_4 fixed as boundary conditions (the dark curve is the initial condition).

This result gives the explicit timestepping stability criterion as $\mathrm{Fo}_M \leq \frac{1}{2}$, and restricts the choice of Δx and Δt. Written in terms of Δt, this criterion is:

$$\Delta t \leq \frac{\Delta x^2}{2\nu}. \quad (2.9)$$

Note that in two dimensions with a square grid, there are four neighbors, and the criterion becomes $\mathrm{Fo}_M \leq \frac{1}{4}$; in three dimensions, $\mathrm{Fo}_M \leq \frac{1}{6}$. Equation 2.9 implies that if one makes the spatial discretization twice as fine (cutting Δx in half), then the timestep must be reduced by a factor of four, requiring eight times the computational work to simulate the same amount of total time.

3.2.2 Implicit Timestepping

Explicit timestepping is a simple algorithm for doing timestepping. Unfortunately, the explicit stability criterion places a very strict limit on the timestep size, making computation very expensive for any decent mesh spacing, but there are methods which can use much much bigger timesteps.
First, explicit timestepping extrapolates the spatial derivatives at the present time forward, resulting in an instability. So why not calculate the future value and use that to interpolate backward? This is the implicit timestepping algorithm, also known as backward Euler time integration, and it does not have the instability associated with the explicit algorithm.
To revisit the example of finite difference simulation of heat conduction, the explicit discretization looks like:

$$\frac{T_{i,n+1} - T_{i,n}}{\Delta t} = \nu \frac{T_{i-1,n} - 2T_{i,n} + T_{i+1,n}}{\Delta x^2} + f_i, \tag{2.10}$$

where f_i is the normalized source term $s_i/\rho c$ at grid point i. With implicit finite differencing, the time indices on the right side change:

$$\frac{T_{i,n+1} - T_{i,n}}{\Delta t} = \nu \frac{T_{i-1,n+1} - 2T_{i,n+1} + T_{i+1,n+1}}{\Delta x^2} + f_i. \tag{2.11}$$

Of course, the new temperatures in timestep $n+1$ are unknown, so how do we calculate these derivatives? The answer is that we don't, explicitly, but we use these as a set of simultaneous equations which we can solve using linear algebra. For the equation above, we can multiply by Δt and rearrange to give:

$$-\frac{\nu \Delta t}{\Delta x^2} T_{i-1,n+1} + \left(1 + 2\frac{\nu \Delta t}{\Delta x^2}\right) T_{i,n+1} - \frac{\nu \Delta t}{\Delta x^2} T_{i+1,n+1} = f_i \Delta t + T_{i,n}. \tag{2.12}$$

This is the equation for one specific interior node; using the mesh Fourier number definition $\mathrm{Fo}_M = \nu \Delta t / \Delta x^2$, and setting temperature boundary conditions at x_0 and x_4 to $T_{0,BC}$ and $T_{4,BC}$ respectively, we can write this equation for an aggregate of nodes:

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ -\mathrm{Fo}_M & 1+2\mathrm{Fo}_M & -\mathrm{Fo}_M & 0 & 0 \\ 0 & -\mathrm{Fo}_M & 1+2\mathrm{Fo}_M & -\mathrm{Fo}_M & 0 \\ 0 & 0 & -\mathrm{Fo}_M & 1+2\mathrm{Fo}_M & -\mathrm{Fo}_M \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} T_{0,n+1} \\ T_{1,n+1} \\ T_{2,n+1} \\ T_{3,n+1} \\ T_{4,n+1} \end{pmatrix} = \begin{pmatrix} T_{0,BC} \\ f_1 \Delta t + T_{1,n} \\ f_2 \Delta t + T_{2,n} \\ f_3 \Delta t + T_{3,n} \\ T_{4,BC} \end{pmatrix}. \tag{2.13}$$

This reduces our difference equations to a simple linear system which we can solve using linear algebra techniques, such as multiplying both sides by the inverse of the matrix on the left.
Implicit timestepping is unconditionally stable. If we take the limit as Δt goes to infinity, we can divide equations 2.12 and 2.13 by the mesh Fourier number, so the 1s in the matrix diagonal and the $T_{i,n}$ on the right side vanish, and we obtain the steady-state result. The downside is that it is considerably more complex to implement than explicit timestepping.
Furthermore, this approach leads to errors on the order of the timestep size. But it is not hard to obtain quadratic or higher accuracy using both explicit and implicit time stepping algorithms known as Adams-Bashforth and Adams-Moulton methods, of which the next section presents one example.

3.2.3 Semi-Implicit Time Integration

Common to explicit and implicit timestepping is the assumption of constant time derivative throughout the timestep. If we acknowledge that the value of $\partial T/\partial t$ may be changing within the timestep, we can achieve better accuracy.

The simplest such approach, known as semi-implicit or Crank-Nicholson timestepping, involves linear interpolation of $\partial T/\partial t$ between timesteps n and $n+1$. Another way to look at this is that it estimates the value of $\partial T/\partial t$ at timestep $n+\frac{1}{2}$. Either way, for the heat equation, we average the right hand sides of equations 2.10 and 2.11 to give:

$$\frac{T_{i,n+1}-T_{i,n}}{\Delta t}=\frac{\nu}{2}\frac{T_{i-1,n}-2T_{i,n}+T_{i+1,n}+T_{i-1,n+1}-2T_{i,n+1}+T_{i+1,n+1}}{\Delta x^2}+\frac{f_{i,n}+f_{i,n+1}}{2}. \tag{2.14}$$

By treating the derivative as linear over the timestep, the error becomes second-order in the timestep size (that is, proportional to Δt^2). This is considerably more accurate than explicit or implicit timestepping, while more stable than explicit timestepping. However, because this average includes the previous timestep, this method is not unconditionally stable (as implicit timestepping is).

3.2.4 Adams-Bashforth

If first order accuracy of explicit and implicit methods are good (like straight Riemann integration), and second order semi-implicit is better (like trapezoid rule integration), why not go on to Simpson's rule for third or higher order discretizations? There are three approaches to accomplishing high-order accuracy in timestepping: an explicit method called Adams-Bashforth, an implicit method called Adams-Moulton, and a very efficient method with multiple function evaluations per timestep called Runge-Kutta.

This explicit higher-order method uses previous timesteps to extrapolate the $\partial T/\partial t$ curve into the next timestep. The simplest such method is just explicit timestepping, which is first-order in time, and can be expressed as:

$$u_{n+1}=u_n+\Delta t f(u_n), \tag{2.15}$$

where u is the unknown variable or set of variables (such as the values of temperature T_i), and

$$\frac{\partial u}{\partial t}=f(u). \tag{2.16}$$

This is analogous to equation 2.7. For higher-order accuracy, one can use previous timesteps, *e.g.* we can use this and the previous timestep to extrapolate a linear function of $\partial T/\partial t$ to achieve quadratic accuracy like semi-implicit timestepping; with uniform Δt, this is written generally as:

$$u_{n+1}=u_n+\Delta t\left[\frac{3}{2}f(u_n)-\frac{1}{2}f(u_{n-1})\right]. \tag{2.17}$$

In general, for polynomial order p:

$$u_{n+1}=u_n+\Delta t\sum_{j=0}^{p-1}a_{p,j}f(u_{n-j}), \tag{2.18}$$

where $\sum a_{p,j}=1$ for any p. This has the advantages of being explicit and solvable without simultaneous equations, while also very accurate, and the function $f(u)$ need only be evaluated once per timestep. The disadvantage is that like explicit/forward Euler time stepping, this method is only stable for suitably small timesteps.

3.2.5 Adams-Moulton

This implicit cousin of Adams-Bashforth time integration interpolates from previous timesteps and the next to provide a polynomial fit estimating $\partial T/\partial t$. Again, the first-order accurate version is the same as implicit/backward Euler time stepping, and the second-order accurate version is identical to semi-implicit/Crank-Nicholson. For a third-order accurate version, we interpolate two old timesteps and the new one to give a quadratic polynomial estimate of $\partial T/\partial t$, which integrates to give third-order accuracy. With uniform Δt, this is written:

$$u_{n+1} = u_n + \Delta t \left[\frac{5}{12} f(u_{n+1}) + \frac{8}{12} f(u_n) - \frac{1}{12} f(u_{n-1}) \right]. \tag{2.19}$$

In general for polynomial order p:

$$u_{n+1} = u_n + \Delta t \sum_{j=0}^{p-1} a_{p,j} f(u_{n-j+1}) \tag{2.20}$$

This is more accurate and more stable than Adams-Bashforth, but as with the comparison between implicit and explicit timestepping, is harder to implement. There is one (non-)linear solution per timestep, with as many function evaluations as necessary to solve the system.

3.2.6 Runge-Kutta Integration

Runge-Kutta integration achieves high-order accuracy by evaluating the function at multiple points within each timestep. Each function evaluation requires no simultaneous equation solution, giving it this advantage over implicit and Adams-Moulton methods. Though a general method, the most often used version is fourth-order accurate, which looks like::

$$u_{n+1} = u_n + \frac{\Delta t}{6}(k_1 + 2k_2 + 2k_3 + k_4), \tag{2.21}$$

where:

$$\begin{aligned} k_1 &= f(u_n), & (2.22)\\ k_2 &= f\left(u_n + \frac{\Delta t}{2} k_1\right), & (2.23)\\ k_3 &= f\left(u_n + \frac{\Delta t}{2} k_2\right), & (2.24)\\ k_4 &= f(u_n + \Delta t k_3). & (2.25) \end{aligned}$$

This takes more time than explicit or Adams-Bashforth methods, particularly if function evaluations are expensive (which they are not for the heat equation), but doesn't require solution of any simultaneous equations.

3.3 Enthalpy Method

When a phase boundary is present in a system, it is necessary to change the formulation somewhat to account for it. Fortunately, for liquid-solid systems there is a relatively simple method for doing

this called the *enthalpy method*, which is based on tracking the enthalpy change at the liquid-solid interface.
The enthalpy uses the relation $\Delta H = \rho c_p \Delta T$, where c_p is the constant-pressure heat capacity. The 1-D enthalpy conservation equation goes:

$$\frac{\partial H}{\partial t} = k\frac{\partial^2 T}{\partial x^2}. \tag{3.26}$$

This is very similar to equation 2.1.
In the enthalpy method, we turn that differential equation directly into a difference equation; with explicit timestepping, this looks like:

$$\begin{aligned} \frac{H_{i,n+1} - H_{i,n}}{\Delta t} &= \frac{\left[k\frac{\partial T}{\partial x}\right]_{i+\frac{1}{2},n} - \left[k\frac{\partial T}{\partial x}\right]_{i-\frac{1}{2},n}}{\Delta x} + s \\ &= \frac{k_{i+\frac{1}{2},n}(T_{i+1,n} - T_{i,n}) - k_{i-\frac{1}{2},n}(T_{i,n} - T_{i-1,n})}{\Delta x^2} + s. \end{aligned} \tag{3.27}$$

Keeping the $k_{i+\frac{1}{2},n}$ and $k_{i-\frac{1}{2},n}$ distinct is necessary because the two sides may be in the different phases with potentially quite different conductivities.
A basic implementation of this stores two fields: H and T. In each timestep, the new H values are calculated from the old H and the old neighboring T values. The new T values are then calculated from the new H values using the inverse of the $H(T)$ function.[1]
This "basic implementation" in one dimension is straightforward to insert into a spreadsheet, as provided for use with Problem Set 2 part 1. It is interesting to watch this simulation in action: during freezing, the moving liquid-solid interface appears to be "pinned" at a spatial location as the enthalpy decreases with no corresponding change in the temperature.
This explicit timestepping formulation (in equation 3.27) will have a stability criterion for the same reason as the non-enthalpy method, and in fact, if the two phases have the same properties, the stability criterion will be the same as equation 2.9. If the two phases have different properties, then the one with the larger ν will have the smaller critical Δt, and choosing that smaller Δt will satisfy the stability criterion in both phases. (Note that at the interface itself, the heat capacity is essentially infinite, so $\nu = 0$ and it is stable for any timestep size.)

3.4 Finite Volume Approach to Boundary Conditions

The finite volume approach is slightly different from finite differences, and I like it because of its elegant approach to presenting boundary conditions. Rather than thinking of the spatial discretization as a set of points, the finite volume approach considers a set of regions, or elements, and considers the average temperature or enthalpy in each element. The heat conservation equation in the enthalpy method can then be written as:

$$V\frac{\partial H}{\partial T} = -\sum A_i q_i + Vs, \tag{4.28}$$

[1] One may also avoid storing the T values by calculating those of the previous timestep from H on the fly (though this makes post-processing the resulting data less straightforward). Although one can store only H and calculate T from that, in many cases one may *not* store only T and calculate H because H is not a unique function of T, *e.g.* in a pure material at the melting temperature, H can not be determined.

where V is the volume of the element (length in one dimension), the A_i are the areas of its faces (one in one dimension), and the q_i are the outward normal fluxes through those faces. At an interior point or volume, this and the finite difference method both give the same results, expressed in equations 2.5 and 3.27.

Based on this, one can very easily express a heat flux boundary condition, or a mixed condition where flux is a function of temperature. A common boundary condition of that type involves a *heat transfer coefficient* labeled as h:

$$\vec{q} \cdot \hat{n} = h(T - T_{env}), \tag{4.29}$$

where T_{env} is an environment temperature. This is often known as a convective boundary condition, based on boundary layer theory of transport in a fluid adjacent to a solid. These types of boundary conditions are straightforward to implement in finite volumes. Note however that this leads to a new stability criterion at the interface, which will put a limit on Δt similar to that in equation 2.9; derivation of this expression is left as an exercise to the reader.

What is not straightforward with this approach is setting a surface temperature. To do this requires setting the flux such that the temperature at the outside face is equal to the desired temperature. One can think of this in terms of a "virtual volume" outside the domain where the temperature is reflected through the desired surface temperature.

Referring again to the spreadsheet, its two boundary conditions are zero flux and convective, making this finite volume approach the logical one. You can see the finite volume implementation in the enthalpy elements on the left, and its boundary conditions on each side, but note that the widths of the boundary elements are half of those of the interior elements. The `castabox` software, also on the website, also uses flux boundary conditions, based on both convective and radiative fluxes (both on the top surface, convective everywhere else). (`castabox` also has an interesting calculation for the stability criterion in three dimensions.)

Finally, note that the finite volume approach can be considered a "zero-order finite element" approach, whose first-order cousin will be introduced in the next chapter, with higher-order variants coming later.

Chapter 4

Weighted Residual Approach to Finite Elements

The weighted residual approach can be used to compute the approximate solution to any partial differential equation, or system of equations, regardless of whether it can be expressed as the minimum of a functional. This approach, also called the Galerkin approach, will be used later to discuss modeling of mechanics and fluid flow, and also forms the basis of the Boundary Element Method (BEM) – and in a sense, the Quantum section of the course.

4.1 Finite Element Discretization

In the finite difference section, we discussed the finite volume approach, in which we consider each cell to have a single average value of temperature (and/or enthalpy). The approximation of the field variable is thus a set of zero-order steps, and the difference between this approximation and the real temperature has errors on the order of the mesh spacing. In finite elements, shapefunctions are piecewise-linear, resulting in higher accuracy; just as the trapezoid rule for integration has errors on the order of the grid spacing squared.
In both finite volume and linear finite element approaches, we consider the approximation of a field variable, such as temperature, using the sum:

$$T \simeq \tilde{T} = \sum_{i=1}^{N} T_i N_i(\vec{x}), \tag{1.1}$$

where $N_i(\vec{x})$ is a *shapefunction* which has the property:

$$N_i(\vec{x}) = \begin{cases} 1, \vec{x} = \vec{x}_i \\ 0, \vec{x} = \vec{x}_{j \neq i} \end{cases}, \tag{1.2}$$

where $\vec{x}_i$ and $\vec{x}_j$ are the *node points*, or gridpoints, in the simulation *mesh*. In one dimension, zero-order elements (finite volume) have a single "gridpoint" on each element, and linear elements have gridpoints on each end.
In two or more dimensions, it is helpful to transform each element to local coordinates, typically written as $\vec{\xi}$ with coordinates ξ, η (and in three dimensions ζ). For example, linear triangle elements

often use a local coordinate system based on the unit right triangle. Then in the local coordinates with gridpoints 1, 2, and 3 at the origin, (1,0) and (0,1) respectively, we have:

$$\begin{aligned} N_1 &= 1 - \xi - \eta, \\ N_2 &= \xi, \\ N_3 &= \eta. \end{aligned} \tag{1.3}$$

We can transform these coordinates back into the reference frame using something like equation 1.1:

$$\vec{x} = \sum_{i=1}^{3} \vec{x}_i N_i(\vec{\xi}), \tag{1.4}$$

for example, if $\vec{\xi} = (1, 0)$, then $N_1(\vec{\xi}) = 0$, $N_2(\vec{\xi}) = 1$, $N_3(\vec{\xi}) = 0$, so $\vec{x} = \vec{x}_2$. Likewise:

$$\vec{\xi} = \left(\frac{1}{3}, \frac{1}{3}\right) \Rightarrow N_1(\vec{\xi}) = N_2(\vec{\xi}) = N_3(\vec{\xi}) = \frac{1}{3}, \tag{1.5}$$

and $\vec{x}$ is just the average of the three corners, which is the centroid of the element.
It is then natural to extend this to second-order elements by considering elements with three gridpoints in each direction: one on each end and one in the middle. In quadratic (and higher-order) square elements we typically use local element coordinates $\xi \in [-1, 1], \eta \in [-1, 1]$, so the shapefunctions are products of $\frac{1}{2}\xi(\xi - 1)$, $1 - \xi^2$ and $\frac{1}{2}\xi(\xi + 1)$. For example, the fourth shapefunction in a quadratic square element (3 nodes $\times$ 3 nodes), corresponding to the fourth node where $\xi = -1, \eta = 0$, is

$$N_4(\vec{\xi}) = \frac{1}{2}\xi(\xi + 1)(1 - \eta^2).$$

4.1.1 Galerkin Approach to Finite Elements

Thus far, we can estimate the field variable and provide a coordinate transformation from elemental to real coordinates. To set up the finite element calculation itself, we can either use the variational formulation discussed by Franz Ulm which minimizes a quantity such as overall energy, or if we have the equation to solve, then the Galerkin weighted-residual approach can be simpler in some ways. For steady-state heat conduction where the partial differential equation is $\nabla^2 T = 0$, this approach defines a residual function as:

$$R_i(T_1, T_2, ..., T_N) = \int \phi_i(\vec{x}) \nabla^2 \tilde{T} dA, \tag{1.6}$$

where ϕ_i is a *weighting function*. If we solve the simultaneous equations to set all of the R_i functions to zero, then we will have a set of temperatures $T_1, ..., T_N$ which comprise an approximate solution to the equation $\nabla^2 T = 0$.
Right away we see there is a problem, because at the element boundaries, the shapefunction derivative $\nabla\phi_i$ is not continuous, so its second derivative $\nabla^2\phi_i$ is not integrable; the same goes for the second derivative of $\tilde{T}$ since that is expressed in terms of the shapefunctions. We can solve this problem using integration by parts:

$$R_i = \int \nabla \cdot (\phi_i \nabla \tilde{T}) dA - \int \nabla\phi_i \cdot \nabla \tilde{T} dA. \tag{1.7}$$

This integral is one we can do with the existing shapefunctions. In typical finite element simulations, we use those shapefunctions N_i for the weighting functions ϕ_i, but that will not be true of *boundary elements*, which will be discussed presently.
To do the overall integral, we can just add the integrals on each element. But to do the element integrals in local element coordinates, we must be able to write dA in terms of $d\xi$ and $d\eta$, which we can do using the cross product of the coordinate gradients from equation 1.4:

$$dA = \begin{vmatrix} \frac{\partial x}{\partial \xi} & \frac{\partial y}{\partial \xi} \\ \frac{\partial x}{\partial \eta} & \frac{\partial y}{\partial \eta} \end{vmatrix} d\xi d\eta. \tag{1.8}$$

This matrix is called the *Jacobian*, and its determinant represents the scaling between the differential areas in the two coordinate systems. In three dimensions, the determinant of the Jacobian corresponds to the triple product of the gradients, and likewise scales the differential volumes.
The integration in equation 1.6 is typically performed using *Gaussian integration*, in which the integral is expressed as a weighted sum of function values at different points:

$$\int f(x)dx = \sum w_i f(x_i), \tag{1.9}$$

where w_i are the integration weights and x_i the integration points. Integrating over an interval with N points allows us to perfectly integrate a $2N+1$-order polynomial, but the math for doing this is beyond the scope of this lecture. I refer you to *Numerical Recipies* for further details.
This all works well for quadratic elements, but for higher-order elements, we need to put the nodes not evenly-spaced, but bunched closer toward the ends. The optimal distribution is at the *Gauss-Lobatto* integration points.
So we can now do great finite-element calculations with arbitrarily high accuracy, as many as ten decimal places if we please. Unfortunately, the constitutive equations and parameters (*e.g.* thermal conductivity) are typically only known to a couple of decimal places, if that. So for the rare calculation where the equation must be solved exactly, high-order polynomial elements, sometimes referred to as "spectral elements", are useful; for most thermal, mechanics and fluids calculations, we just use linear or quadratic elements.

4.2 Green's Functions and Boundary Elements

Green's functions are tools used to solve (partial) differential equations of the form:

$$Lu + f = 0, \tag{2.10}$$

where L is a linear operator, f is a "source" function, and u is the unknown field variable. For example, back to the (steady-state) heat equation, L becomes $k\nabla^2$ (or $\nabla \cdot (k\nabla)$ for non-uniform k) and f becomes $\dot{q}$, bringing back the familiar equation:

$$\nabla \cdot (k\nabla T) + \dot{q} = 0. \tag{2.11}$$

The Green's function $u*$ is the solution to the equation:

$$Lu* + \delta(\vec{x} - \vec{x}') = 0, \tag{2.12}$$

where $\vec{x}'$ is a reference point, and the delta function is defined such that:

$$\int_\Omega \delta(\vec{x} - \vec{x}')d\vec{x} = \begin{cases} 1, \vec{x}' \in \Omega, \\ 0 \text{ otherwise.} \end{cases} \tag{2.13}$$

The Green's function can be used by itself, or in the boundary element method. Both of those uses will be explored further in lecture.

4.3 Fourier Series Methods

This roughly corresponds to the "modal analysis" as described by Franz Ulm, as opposed to "direct time integration", in that the matrix describing the system is diagonalized due to the weak interactions between the various waves.
This is most straightforwardly illustrated in Fourier series solution of the heat equation.
Unfortunately, due to insufficient time this topic was not covered this year. A future version of these notes may include a complete section on fourier methods.

Chapter 5

Linear Elasticity

This lecture note continues our investigation of continuum modeling and simulation into linear ealstic problems. We start with an interesting analogy that can be made between the 1-D heat diffusion problem and 1-D elasticity. We then have a closer look on the physics of the problem, and develop the basis of the Finite Element Method for elasticity problems: the principle of virtual work. By way of application we investigate the water filling of a dam by means of finite element simulations.

5.1 From Heat Diffusion to Elasticity Problems

How to start?

5.1.1 1-D Heat Diffusion – 1-D Elasticity Analogy

There is an interesting analogy to be made between a steady state 1-D heat diffusion problem and a 1-D elasticity problem of a deforming truss (see Figure 1.1):

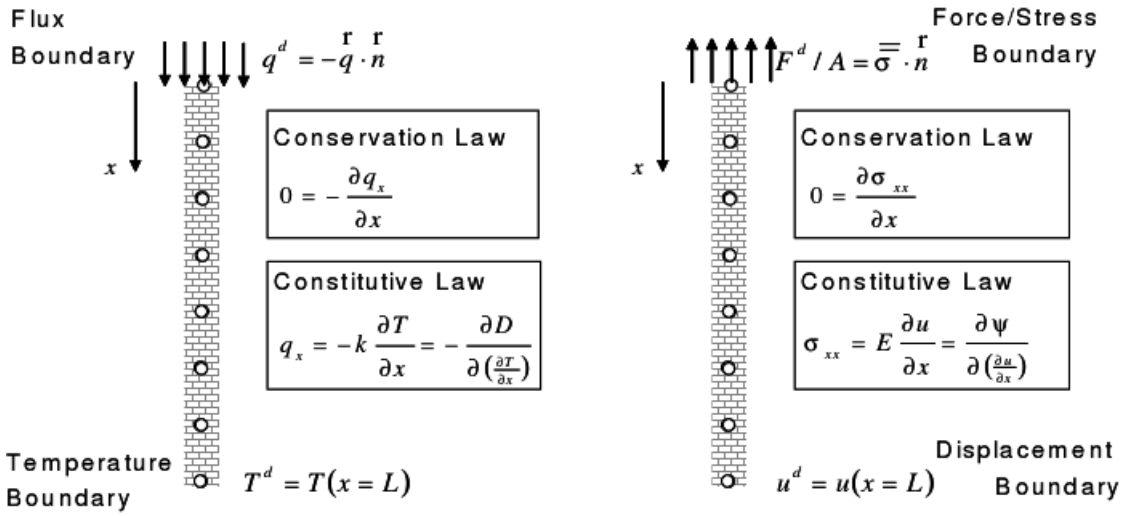

Figure 1.1: Analogy: 1-D heat diffusion vs. 1-D Truss Elasticity.

1. Boundary Conditions: The boundary conditions in the heat problem are either temperature or heat flux boundary conditions; and in the 1-D truss problem they are either displacement boundary conditions or a force boundary condition;

$$x = 0; x = L : \left\{ \begin{array}{c} \text{either} : T^d = T\left(0/L\right) \Rightarrow u^d = u\left(0/L\right) \\ \text{or} : q^d = -q_x \qquad \Rightarrow F^d/S = \sigma \end{array} \right\} \tag{1.1}$$

where u^d is a prescribed displacement, F^d a prescribed force (positive in tension, whence the difference in sign wrt the heat problem), S is the surface, so that σ is a force per unit surface, that is stress.

2. Conservation Law: The conservation law in the heat problem is the energy conservation, and in the 1-D truss problem it is the conservation of the momentum:

$$0 < x < L : 0 = -\frac{\partial q_x}{\partial x} \Rightarrow 0 = \frac{\partial \sigma}{\partial x} \tag{1.2}$$

3. Constitutive Law: The constitutive law in the heat problem is Fourier's Law, linking the heat flux to the temperature gradient; and the constitutive law in the 1-D truss problem is a link between the stress and the displacement gradient, which is known as Hooke's Law:

$$q_x = -k\frac{\partial T}{\partial x} \Rightarrow \sigma = E\frac{\partial u}{\partial x} \tag{1.3}$$

E is the Young's modulus. The displacement gradient is called strain $\varepsilon = \frac{\partial u}{\partial x}$. Alternatively, the constitutive analogy can be made using potential functions:

$$q_x = -\frac{\partial \mathcal{D}}{\partial\left(\frac{\partial T}{\partial x}\right)} \Rightarrow \sigma = \frac{\partial \psi}{\partial\left(\frac{\partial u}{\partial x}\right)} = \frac{\partial \psi}{\partial \varepsilon} \tag{1.4}$$

where ψ is the so-called free energy (or Helmholtz energy). For instance for a linear elastic material, ψ reads:

$$\psi = \frac{1}{2}E\left(\frac{\partial u}{\partial x}\right)^2 = \frac{1}{2}E\varepsilon^2 \tag{1.5}$$

It is readily understood that ψ is a convex function of its argument (the displacement gradient), so that:

$$\frac{\partial \psi}{\partial \varepsilon}\left(\varepsilon^* - \varepsilon\right) \leq \psi\left(\varepsilon^*\right) - \psi\left(\varepsilon\right) \tag{1.6}$$

For a linear elastic material, strict convexity has an obvious physical meaning: the stiffness E must be greater than zero,

$$E = \frac{\partial^2 \psi}{\partial \varepsilon^2} > 0 \tag{1.7}$$

5.1.2 3-D Extension

The heat diffusion – elasticity analogy holds also for the 3-D situation, if we make the link in between the 3-D counterparts of the 1-D quantities. This analogy is given by:

$$\begin{array}{rcl} T & \rightarrow & \overrightarrow{u} = \underline{u} \\ \overrightarrow{q} = \underline{q} & \rightarrow & \underline{\underline{\sigma}} \\ -\overrightarrow{\bigtriangledown T} = \lim gradT & \rightarrow & \underline{\underline{\varepsilon}} = \bigtriangledown^s \overrightarrow{u} = \frac{1}{2}\left(\lim grad\underline{u} + (\lim grad\underline{u})^T\right) \\ \text{Fourier's law} : \underline{q} = -\dfrac{\partial \mathcal{D}}{\partial\left(\lim gradT\right)} & \rightarrow & \text{Hooke's law} : \underline{\underline{\sigma}} = \dfrac{\partial \psi}{\partial \underline{\underline{\varepsilon}}} \\ \text{Heat balance} : \lim \text{div}\underline{q} = 0 & \rightarrow & \text{Momentum balance} : \lim \text{div}\underline{\underline{\sigma}} = 0 \end{array} \tag{1.8}$$

There are some important differences to be noted when we go from 1-D to 3-D:

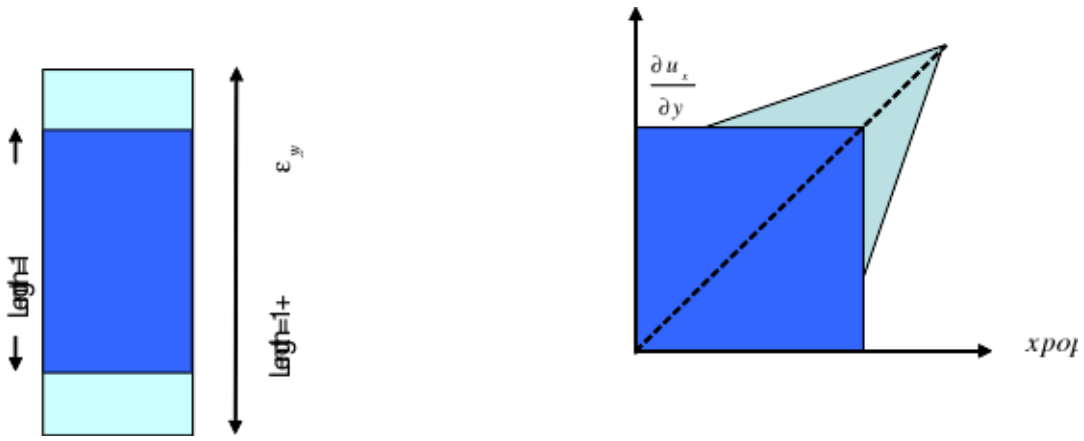

Figure 1.2: Deformation Modes: Stretching (left), Distortion (right).

1. The elasticity analogous of the temperature T, which is a scalar, is the displacement *vector*, which in 3-D has three components. As a consequence, the quantities that intervene in the elasticity problem are all of a higher spatially order. For instance, the analogous of the temperature gradient, which is a vector, is the symmetric part of the displacement gradient $\bigtriangledown^s \overrightarrow{u}$. Since the displacement is a vector (or 1st order tensor), the gradient of this vector is a 3×3 strain matrix (or 2nd order tensor):

$$\underline{\underline{\varepsilon}} = \begin{bmatrix} \varepsilon_{xx} = \dfrac{\partial u_x}{\partial x} & \varepsilon_{xy} = \dfrac{1}{2}\left[\dfrac{\partial u_x}{\partial y} + \dfrac{\partial u_y}{\partial x}\right] & \varepsilon_{xz} = \dfrac{1}{2}\left[\dfrac{\partial u_x}{\partial z} + \dfrac{\partial u_z}{\partial x}\right] \\ & \varepsilon^*_{yy} = \dfrac{\partial u_y}{\partial y} & \varepsilon_{yz} = \dfrac{1}{2}\left[\dfrac{\partial u_y}{\partial z} + \dfrac{\partial u_z}{\partial y}\right] \\ \text{sym} & & \varepsilon^*_{zz} = \dfrac{\partial u_z}{\partial y} \end{bmatrix} \tag{1.9}$$

The diagonal terms of this matrix represent relative length variations, and the out-of-diagonal terms are shear strains representing distortions. Their geometrical significance is sketched in figure 1.2.

2. The analogous of the heat flux *vector* $\overrightarrow{q}$, is a symmetric 3×3 stress matrix (or 2nd order tensor), $\underline{\underline{\sigma}}$:

$$\underline{\underline{\sigma}} = \begin{bmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ & \sigma_{yy} & \sigma_{yz} \\ \text{sym} & & \sigma_{zz} \end{bmatrix} \tag{1.10}$$

The stress is at equilibrium if it satisfies the momentum balance equation:

$$\lim div \underline{\underline{\sigma}} + \rho \overrightarrow{g} = 0 \tag{1.11}$$

where ρ is the mass density and $\overrightarrow{g}$ is the earth acceleration vector. For instance, water at hydrostatic equilibrium has as stress:

$$\underline{\underline{\sigma}} = -p \begin{bmatrix} 1 & 0 & 0 \\ & 1 & 0 \\ \text{sym} & & 1 \end{bmatrix} \tag{1.12}$$

where p is the fluid pressure. The momentum balance of water at rest thus reads:

$$-\begin{pmatrix} \dfrac{\partial}{\partial x} \\ \dfrac{\partial}{\partial y} \\ \dfrac{\partial}{\partial z} \end{pmatrix} p + \rho_w \begin{pmatrix} g_x \\ g_y \\ g_z \end{pmatrix} = 0 \tag{1.13}$$

where (g_x, g_y, g_z) are the components of the acceleration vector. Solids (or moving fluids) have in addition shear stresses, and the momentum balance (1.11) reads:

$$\frac{\partial \sigma_{xx}}{\partial x} + \frac{\partial \sigma_{xy}}{\partial y} + \frac{\partial \sigma_{xz}}{\partial z} + \rho g_x = 0 \tag{1.14}$$

$$\frac{\partial \sigma_{xy}}{\partial x} + \frac{\partial \sigma_{yy}}{\partial y} + \frac{\partial \sigma_{yz}}{\partial z} + \rho g_y = 0 \tag{1.15}$$

$$\frac{\partial \sigma_{xz}}{\partial x} + \frac{\partial \sigma_{yz}}{\partial y} + \frac{\partial \sigma_{zz}}{\partial z} + \rho g_z = 0 \tag{1.16}$$

3. The constitutive analogous of Fourier's Law is Hooke's Law. Similarly to Fourier's law, which provides a link between the heat flux vector and the temperature gradient, Hooke's law provides a link between the stress matrix and the strain matrix. This link needs to account for the possible deformation and energy modes the material can undergo. Here we restrict ourselves to the isotropic situation: Isotropy means that the behavior has no privileged direction. In this isotropic case, we need to consider the two deformation modes shown in figure 1.2, by means of two elasticity constants:

$$\psi = \frac{\lambda}{2} \left(\lim tr \underline{\underline{\varepsilon}} \right)^2 + \mu \lim tr \left(\underline{\underline{\varepsilon}} \cdot \underline{\underline{\varepsilon}} \right) \tag{1.17}$$

where $\lim tr \underline{\underline{\varepsilon}} = \varepsilon_{xx} + \varepsilon_{yy} + \varepsilon_{zz}$ is the sum of the diagonal terms in (1.9), while $\lim tr \left(\underline{\underline{\varepsilon}} \cdot \underline{\underline{\varepsilon}} \right)$ is the sum of the diagonal terms of the matrix product $\underline{\underline{\varepsilon}} \cdot \underline{\underline{\varepsilon}}$. λ and μ are called Lamé constants, and are related to the Young's modulus by:

$$\lambda = \frac{\nu E}{(1+\nu)(1-2\nu)}; \ \mu = \frac{E}{2(1+\nu)} \tag{1.18}$$

where $-1 \leq \nu \leq 0.5$ is the Poisson's ratio, a dimensionless number which captures essentially the volumetric versus the shear deformability of an isotropic material. For instance, $\nu = 0.5$ corresponds to an incompressible material, $\lim tr\underline{\underline{\varepsilon}} = 0$, for which $\lambda \to \infty$. Using the free energy expression (1.17), the stress-strain (or displacement gradient relation) is readily obtained:

$$\underline{\underline{\sigma}} = \frac{\partial \psi}{\partial \underline{\underline{\varepsilon}}} = \lambda \left(\lim tr\underline{\underline{\varepsilon}} \right) \lim diag\,[1] + 2\mu \underline{\underline{\varepsilon}} \tag{1.19}$$

Here, $\lim diag\,[1]$ stands for a 3×3 diagonal unit matrix.

Provided these (minor, but important) differences the analogy (1.8) is readily put to work.

5.2 The Theorem of Virtual Work

Hold on... what's about the physics?

5.2.1 From the Theorem of Virtual Work in 1-D to Finite Element Formulation

We start with the 1-D truss problem displayed in figure 1.1. The work that is provided from the outside to the system is:

$$W_{ext} = F^d u + F u^d = \int_S \sigma u \; dS \tag{2.20}$$

where F^d stands for prescribed forces and u^d for prescribed displacements at the boundary S, i.e. on $x = 0$ and $x = L$. This work must be equal to the work done in the inside of the truss, which we obtain by application of the divergence theorem (in 1-D):

$$W_{ext} = W_{int} = \int_{x=0}^{x=L} \frac{\partial}{\partial x} \left(\sigma u \right) \, dx = \int_{x=0}^{x=L} \left(u \frac{\partial \sigma}{\partial x} + \sigma \frac{\partial u}{\partial x} \right) \, dx \tag{2.21}$$

The stress satisfies the momentum balance $\frac{\partial \sigma_{xx}}{\partial x} = 0$, so that the equality of the external and the internal work realized by the stress σ_{xx} reads:

$$F^d u + F u^d = \int_{x=0}^{x=L} \sigma \frac{\partial u}{\partial x} \, dx = \int_{x=0}^{x=L} \sigma \varepsilon \, dx \tag{2.22}$$

Expression (2.22) is the 1-D version of the theorem of virtual work. Note that the stress and the displacement gradient are not yet related by a material law of the form (1.3) or (1.4). So we can apply this theorem with any other displacement $u^*(x)$, which satisfies the displacement boundary condition:

$$u^*(L) = u^d \tag{2.23}$$

This yields:

$$F^d u^* + F u^d = \int_{x=0}^{x=L} \sigma \varepsilon^* \, dx \tag{2.24}$$

Subtracting (2.22) from (2.24) yields:

$$F^d\left(u^* - u\right) = \int_{x=0}^{x=L} \sigma\left(\varepsilon^* - \varepsilon\right)\, dx \tag{2.25}$$

It is only at this stage that we introduce the constitutive relation (1.4), which links the stress solution σ with the displacement gradient solution $\frac{\partial u}{\partial x}$, so that the integrand of (2.25) reads:

$$\sigma\left(\varepsilon^* - \varepsilon\right) = \frac{\partial \psi}{\partial \varepsilon}\left(\varepsilon^* - \varepsilon\right) \tag{2.26}$$

We can now make use of the convexity property (1.6) of the free energy ψ on the right hand side of (2.26), and substitute this inequality into (2.25). We so obtain

$$F^d\left(u^* - u\right) \leq \int_{x=0}^{x=L} \left[\psi\left(\varepsilon^*\right) - \psi\left(\varepsilon\right)\right]\, dx \tag{2.27}$$

and after re-arrangement:

$$\int_{x=0}^{x=L} \psi\left(\varepsilon\right)\, dx - F^d u \leq \int_{x=0}^{x=L} \psi\left(\varepsilon^*\right)\, dx - F^d u^* \tag{2.28}$$

The difference between the overall free energy contained in the truss and the work provided by prescribed forces is the potential energy, and is denoted by $\mathcal{E}_{pot}\left(u\right)$. What inequality (2.28) then shows is that among all possible displacement solutions $u^*\left(x\right)$ that satisfy the displacement boundary condition (2.23), the solution $u\left(x\right)$ minimizes the potential energy:

$$\mathcal{E}_{pot}\left(u\right) = \min_{u^* = u^d \lim on S} \left(\int_{x=0}^{x=L} \psi\left(\varepsilon^*\right)\, dx - F^d u^*\right) \tag{2.29}$$

By analogy with the heat diffusion problem, we discretize the continuous displacement field by:

$$u^*\left(x\right) = \sum_{i=0}^{M+1} u_i N_i(x) \tag{2.30}$$

where u_i are nodal displacements, and $N_i(x)$ are the base functions. For a linear elastic material for which $\psi\left(\varepsilon^*\right)$ is given by (1.5), the potential energy reads:

$$\mathcal{E}_{pot}\left(u^*\right) = \int_{x=0}^{x=L} \frac{1}{2} E \left(\frac{\partial u^*}{\partial x}\right)^2 dx - F^d u^* = \sum_{i=1}^{M}\sum_{j=1}^{M} \frac{1}{2} u_i K_{i,j} u_j - F_i^d u_i^* \tag{2.31}$$

where K_{ij} are elements of the stiffness matrix:

$$K_{i,j} = \int_{x=0}^{x=L} E\left(\frac{\partial N_i}{\partial x}\frac{\partial N_j}{\partial x}\right)\, dx \tag{2.32}$$

Then, in order to find the minimum of the potential energy $\mathcal{E}_{pot}\left(u^*\right)$, which comes the closest to the potential energy of the solution $\mathcal{E}_{pot}\left(u\right)$, we minimize (2.31) wrt the displacement degrees of freedom u_i, and obtain a system of M linear equations:

$$\frac{\partial \mathcal{E}_{pot}\left(u^*\right)}{\partial u_i} = 0 \Rightarrow \sum_{j=1}^{M} K_{ij} u_j - F_i^d = 0 \tag{2.33}$$

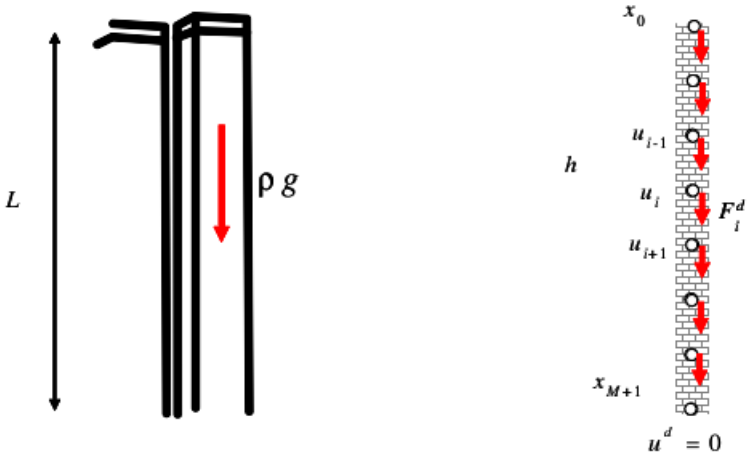

Figure 2.3: Column under self-weight: Architectural view (left) – Engineering Finite Element Model (right).

Example 5.2.1 *By way of application, let us consider a column of length L subjected to its self weight, as sketched in figure 2.3. The column is clamped at $x = L$, and force free at its top. We discretize the column by 3 finite elements of same length $h = L/3$ and employ linear base functions. The stiffness matrix components for the internal nodes $(i = 1, 2)$ are readily obtained from (2.32):*

$$K_{i,i-1} = \int_{x_{i-1}}^{x_i} E\left(-\frac{1}{h^2}\right) dx = -\frac{E}{h} \tag{2.34}$$

$$K_{i,i} = \int_{x_{i-1}}^{x_{i+1}} E\left(\frac{1}{h^2}\right) dx = +\frac{2E}{h} \tag{2.35}$$

$$K_{i,i+1} = \int_{x_i}^{x_{i+1}} E\left(-\frac{1}{h^2}\right) dx = -\frac{E}{h} \tag{2.36}$$

Similarly, the prescribed forces at each node are obtained by integrating the self weight (per unit section):

$$F_0^d = \int_0^{x_1} \rho g N_i(x)\, dx = \frac{\rho g h}{2} \tag{2.37}$$

$$F_i^d = \int_{x_{i-1}}^{x_{i+1}} \rho g N_i(x)\, dx = \rho g h \tag{2.38}$$

where A is the column section, ρ is the mass density, g is the earth acceleration. The equations for the different nodes read:

$$\begin{array}{c} i = 0 : K_{0,0}u_0 + K_{0,1}u_1 - F_0^d = 0 \\ i = 1 : K_{1,0}u_0 + K_{1,1}u_1 + K_{1,2}u_2 - F_1^d = 0 \\ i = 2 : K_{2,1}u_1 + K_{2,2}u_2 + K_{2,3}u_3 - F_2^d = 0 \end{array} \tag{2.39}$$

If we note that the displacement $u^*(x = L) = u_3 = 0$, the system of equation (2.31) to be solved

reads:

$$\frac{E}{h}\begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}\begin{pmatrix} u_0 \\ u_1 \\ u_2 \end{pmatrix} - \frac{\rho g h}{2}\begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \tag{2.40}$$

Using standard tools of solving linear system of equations, we obtain:

$$\begin{pmatrix} u_0 \\ u_1 \\ u_2 \end{pmatrix} = \frac{\rho g h^2}{2E}\begin{pmatrix} 9 \\ 8 \\ 5 \end{pmatrix} = \frac{\rho g L^2}{18E}\begin{pmatrix} 9 \\ 8 \\ 5 \end{pmatrix} \tag{2.41}$$

It is instructive to see how the solution improves as we double the number of nodes, so that $h' = h/2 = L/6$*,with the boundary condition* $u^*(x = L) = u_6 = 0$*. The system to be solved then reads:*

$$\frac{E}{h'}\begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & -1 & 2 \end{bmatrix}\begin{pmatrix} u'_0 \\ u'_1 \\ u'_2 \\ u'_3 \\ u'_4 \\ u'_5 \end{pmatrix} - \frac{\rho g h'}{2}\begin{pmatrix} 1 \\ 2 \\ 2 \\ 2 \\ 2 \\ 2 \end{pmatrix} = 0 \tag{2.42}$$

The solution reads:

$$\begin{pmatrix} u'_0 \\ u'_1 \\ u'_2 \\ u'_3 \\ u'_4 \\ u'_5 \end{pmatrix} = \frac{\rho g (h')^2}{2E}\begin{pmatrix} 36 \\ 35 \\ 32 \\ 27 \\ 20 \\ 11 \end{pmatrix} = \frac{\rho g L^2}{72E}\begin{pmatrix} 36 \\ 35 \\ 32 \\ 27 \\ 20 \\ 11 \end{pmatrix} \tag{2.43}$$

The solution so obtained is found not to improve the solution. Finally, a comparison with the exact solution is due. The continuous equation to be solved reads:

$$\frac{\partial \sigma}{\partial x} + \rho g = 0 \tag{2.44}$$

or equivalently using the constitutive law (1.2):

$$E\frac{d^2 u}{dx^2} + \rho g = 0 \tag{2.45}$$

Integration yields the stress solution:

$$\sigma(x) = E\left(u'_0 - \frac{\rho g}{E}x\right)$$

and the displacement solution:

$$u(x) = u_0 + u'_0 x - \frac{\rho g}{2E}x^2 \tag{2.46}$$

where u_0 *and* u'_0 *are two integration constants that need to be solved by boundary conditions. The first boundary condition is that the stress is zero at* $x = 0$*, the second that the displacement is zero at* $x = L$*. It follows:*

$$\sigma(x = 0) = 0 \Rightarrow u'_0 = 0 \tag{2.47}$$

$$u(x = L) = 0 \Rightarrow u_0 = \frac{\rho g L^2}{2E} \tag{2.48}$$

Hence, from a comparison of the FE-solutions and the exact solutions, we conclude that both discretized solutions at their respective nodal points are exact solutions. In fact, two finite element with three nodal points would have been sufficient to capture the parabolic displacement solution. Application: The World Trade Center Towers were roughly $L = 400\text{m}$ *high and* $A = 60 \times 60\text{m}^2$ *wide and had a weight of approximately* $M = 300,000,000\text{kg}$. *The average mass density can be estimated from*

$$\rho \sim \frac{M}{LA} \approx 200 \frac{\text{kg}}{\text{m}^3}$$

The effective stiffness can be estimated by averaging the stiffness of the columns over the floor:

$$E = E_s \frac{A_c}{A}$$

where $E_s = 210\text{GPa}$. *The towers were built as steel tubular structural system with* 4×59 *box columns at the perimeter and* 44 *box columns in the core, thus a total of* 280 *columns. If we estimate that each columns has roughly a section of* 0.20m^2, *an order of magnitude of the column-to-floor section is* $A_c/A \sim 2\%$, *whence an effective stiffness in the vertical direction of* $E \sim 4\text{GPa} = 4,000,000,000\text{Pa}$. *The vertical deformation at the top of the towers should have been on the order of:*

$$u_0 = \frac{200 \times 10 \times 400^2}{2 \times 4,000,000,000} = 0.04\text{m}$$

This is really not much ($u_0/L = 1/10,000$), *so there is not much to be worried when it comes to the vertical deformation under self-weight. The critical design loads for skyscrapers are horizontal loads: winds, earthquakes, etc.*

5.2.2 Theorem of Minimum Potential Energy in 3-D Linear Isotropic Elasticity

What we have seen in 1-D extends to 3-D linear isotropic elasticity problems, if we replace the 1-D scalar quantities by their 3-D counterparts. The 3-D external work rate comprises two terms, the work done by volume forces (such as the gravity force vector $\rho \overrightarrow{g}$, as just seen in the example), and the work by surface traction vectors $\overrightarrow{t} = \underline{\underline{\sigma}} \cdot \overrightarrow{n}$:

$$W_{ext} = \int_V \overrightarrow{u} \cdot \rho \overrightarrow{g} \, dV + \int_S \overrightarrow{u} \cdot \left(\underline{\underline{\sigma}} \cdot \overrightarrow{n} \right) \, dS \tag{2.49}$$

Application of the principle of the divergence theorem to the second term yields the internal work:[1]

$$W_{int} = \int_V \underline{\underline{\sigma}} : \underline{\underline{\varepsilon}} \, dV \tag{2.50}$$

[1]The derivation is not trivial. First of all, application of the divergence theorem reads here:

$$\int_S \overrightarrow{u} \cdot \left(\underline{\underline{\sigma}} \cdot \overrightarrow{n} \right) \, dS = \int_S \lim div \left(\overrightarrow{u} \cdot \underline{\underline{\sigma}} \right) \, dV = \int_S \left[\overrightarrow{u} \cdot \lim div \left(\underline{\underline{\sigma}} \right) + \lim grad \overrightarrow{u} : \underline{\underline{\sigma}} \right] \, dV$$

If we note that $\underline{\underline{\sigma}}$ is symmetric, we have:

$$\lim grad \overrightarrow{u} : \underline{\underline{\sigma}} = \underline{\underline{\varepsilon}} : \underline{\underline{\sigma}} = \underline{\underline{\sigma}} : \underline{\underline{\varepsilon}}$$

In matrix notation, the double product $\underline{\underline{\varepsilon}} : \underline{\underline{\sigma}}$ corresponds to taking the trace of the matrix product $\underline{\underline{\varepsilon}} \cdot \underline{\underline{\sigma}}$.

$$\underline{\underline{\varepsilon}} : \underline{\underline{\sigma}} = \lim tr \left(\underline{\underline{\varepsilon}} \cdot \underline{\underline{\sigma}} \right)$$

The theorem of virtual work then reads:

$$\begin{aligned} W_{ext} &= W_{int} \\ \int_V \overrightarrow{u} \cdot \rho \overrightarrow{g}\, dV + \int_S \overrightarrow{u} \cdot \left(\underline{\underline{\sigma}} \cdot \overrightarrow{n}\right)\, dS &= \int_V \underline{\underline{\sigma}} : \underline{\underline{\varepsilon}}\, dV \end{aligned} \tag{2.51}$$

Following the 1-D approach, we apply the theorem of virtual work with a second displacement field $\overrightarrow{u}^*$ which satisfies the displacement boundary condition:

$$\int_V \overrightarrow{u}^* \cdot \rho \overrightarrow{g}\, dV + \int_S \overrightarrow{u}^* \cdot \left(\underline{\underline{\sigma}} \cdot \overrightarrow{n}\right)\, dS = \int_V \underline{\underline{\sigma}} : \underline{\underline{\varepsilon}}^*\, dV \tag{2.52}$$

Taking the difference of (2.51) and (2.52) yields the 3-D generalization of (2.25):

$$\int_V (\overrightarrow{u}^* - \overrightarrow{u}) \cdot \rho \overrightarrow{g}\, dV + \int_S (\overrightarrow{u}^* - \overrightarrow{u}) \cdot \overrightarrow{t}^d\, dS = \int_V \underline{\underline{\sigma}} : \left(\underline{\underline{\varepsilon}}^* - \underline{\underline{\varepsilon}}\right) dV \tag{2.53}$$

where $\overrightarrow{t}^d$ is the prescribed surface traction vector. Finally, all what we are left to do is to recognize that the stress $\underline{\underline{\sigma}}$ in (2.53) derives from the free energy ψ (see Eq. (1.19)), and that the free energy is a convex function of its argument $\underline{\underline{\varepsilon}}$; that is analogously to the 1-D situation (1.6):

$$\underline{\underline{\sigma}} : \left(\underline{\underline{\varepsilon}}^* - \underline{\underline{\varepsilon}}\right) = \frac{\partial \psi}{\partial \underline{\underline{\varepsilon}}} : \left(\underline{\underline{\varepsilon}}^* - \underline{\underline{\varepsilon}}\right) \leq \psi\left(\underline{\underline{\varepsilon}}^*\right) - \psi\left(\underline{\underline{\varepsilon}}\right) \tag{2.54}$$

Finally, replacing the r.h.s. of equality (2.54) by the r.h.s. of inequality (2.54), we arrive, after some rearrangements, at the 3-D formulation of the theorem of minimum potential energy (2.29):

$$\mathcal{E}_{pot}\left(\overrightarrow{u}\right) = \min_{\overrightarrow{u}^* = \overrightarrow{u}^d \lim onS} \left(\int_V \psi\left(\underline{\underline{\varepsilon}}^*\right) dV - \left(\int_V \overrightarrow{u}^* \cdot \rho \overrightarrow{g}\, dV + \int_S \overrightarrow{u}^* \cdot \overrightarrow{t}^d\, dS \right) \right) \tag{2.55}$$

The theorem states that among all possible displacement fields $\overrightarrow{u}^*$ that satisfy the displacement boundary conditions, the solution field $\overrightarrow{u}$ minimizes the potential energy. The first term on the right hand side of (2.55) represents the entire free energy which is contained in the system, and the two second terms represent the work provided from the outside through volume and surface forces.

We then obtain with (2.49):

$$W_{ext} = \int_V \overrightarrow{u} \cdot \rho \overrightarrow{g}\, dV + \int_V \left[\overrightarrow{u} \cdot \lim div \underline{\underline{\sigma}} + \lim tr\left(\underline{\underline{\varepsilon}} \cdot \underline{\underline{\sigma}}\right)\right]\, dV$$

We now need to consider the local momentum balance equation, which reads in the presence of body forces (see Eq. (1.11)):

$$\lim div \underline{\underline{\sigma}} = -\rho \overrightarrow{g}$$

The first term in the external work skips out, and we find that:

$$W_{ext} = \int_V \lim tr\left(\underline{\underline{\varepsilon}} \cdot \underline{\underline{\sigma}}\right)\, dV = W_{int}$$

where $\lim tr\left(\underline{\underline{\varepsilon}} \cdot \underline{\underline{\sigma}}\right)$ represents in fact the internal work achieved by the stress $\underline{\underline{\sigma}}$ along the displacement gradient $\underline{\underline{\varepsilon}}$.

5.2.3 3-D Finite Element Implementation

Similarly to the 1-D situation, we approximate the continuous displacement field through a discrete representation:

$$\overrightarrow{u}^* = \sum_{i=0}^{i=M+1} u_i N_i(x_j) \tag{2.56}$$

where u_i represent discrete displacements at nodal points, and $N_i(x_j)$ are the base functions. Analogously to the 1-D case, we arrive at a very similar expression of the potential energy $\mathcal{E}_{pot}(\overrightarrow{u}^*)$ as in the 1-D case:

$$\mathcal{E}_{pot}(\overrightarrow{u}^*) = \frac{1}{2} u_i K_{ij} u_j - F_i^d u_i \tag{2.57}$$

where K_{ij} is still the stiffness matrix, and F_i^d the external force vector which here accounts for both the volume forces and surface forces. As a consequence of the theorem of minimum potential energy (2.55), the problem to be solved reduces to solving a linear system of equations:

$$\frac{\partial \mathcal{E}_{pot}(u_i, u_j)}{\partial u_i} = 0 \Rightarrow K_{ij} u_j - F_i^d = 0 \tag{2.58}$$

We will see an interesting application here below.

5.2.4 Homework Set: Water Filling of a Gravity Dam

We consider a gravity dam of triangular shape (Height H, Top angle α) clamped on its base OA (zero displacement prescribed), see figure 2.4. The down-stream surface AB is stress free. The upstream wall OB is subjected to the water pressure (body force $\rho_w \overrightarrow{g}$) over a height $OB' = mH$, where $m \in [0, 1]$. The water is at rest satisfying hydrostatic equilibrium. The remaining surface $B'B$ is stress free. For the purpose of analysis, we assume that the gravity dam is composed of a linear isotropic elastic material (Lamé constants λ, μ) with volume mass ρ_d. With regard to the dimension of the gravity dam in the $Oz-$direction, the problem can be treated as a plane strain problem with regard to the plane parallel to Oxy.
We want to evaluate the displacement of the dam along the upstream wall OB with water filling $m \in [0, 1]$.

Exercise 5.2.1 ***First Approximation****: We consider a displacement field of the form:*

$$\overrightarrow{u}^* = a \frac{y}{H} \overrightarrow{e}_x + b \frac{y}{H} \overrightarrow{e}_y \tag{2.59}$$

Using the theorem of minimum potential energy, give an approximation of the displacement along OB. Show that the problem can be recast in the format of the finite element method:

$$K_{ij} u_j = F_i^d \tag{2.60}$$

where K_{ij} are the components of the stiffness matrix, u_j are the unknown nodal displacements, and F_i^d are nodal forces. Determine K_{ij}, F_i^d and u_j for the given approximation (2.59) of the displacement field $\overrightarrow{u}^*$ in the gravity dam.

Exercise 5.2.2 ***Finite Element Approximation****: Consider $m = 1$, and $\rho_d = 0$. By means of the finite elements, show that the horizontal displacement solution converges with increasing number of elements. To this end, normalize the obtained displacement at $y = H$ by*

$$\overline{u} = \frac{u_x(x = 0, y = H)}{\left(\frac{\rho_w g H^2}{3\mu \tan\alpha}\right)} \tag{2.61}$$

with μ the shear modulus. Represent your results in form of a graph giving $\overline{u}$ as a function of the number of elements over OB.

Image removed due to copyright reasons.

Figure 2.4: Hydroelectric Dam in Krasnoyarsk, Russia (1967). Left: Photo by Dr. A. Hugentobler; With permission from `http://www.structurae.net`; Right: Engineering Model.

The One-Triangle Finite Element Solution

The system we are interested is the solid dam, subjected to displacement and stress boundary conditions:

1. We need to check that the displacement field (2.59) is compatible with the boundary conditions. This is the case here:

$$\lim on OA : \overrightarrow{u}^* (x, y = 0) = \overrightarrow{u}^d = 0 \tag{2.62}$$

2. We need to determine the surface tractions the upstream water exerts on the surface OB'. The stress in the water is at hydrostatic equilibrium, as defined by (1.13). In the coordinate

system of the problem, the acceleration vector reads $\overrightarrow{g} = -g\,\overrightarrow{e}_y$ or in vector components $(g_x, g_y, g_z) = (0, -g, 0)$, where $g \simeq 10\text{m/s}^2$. The hydrostatic momentum balance equation then reads:

$$-\begin{pmatrix} \dfrac{\partial}{\partial x} \\ \dfrac{\partial}{\partial y} \\ \dfrac{\partial}{\partial z} \end{pmatrix} p - \rho_w \begin{pmatrix} 0 \\ g \\ 0 \end{pmatrix} = 0 \tag{2.63}$$

The first and the last equation show that the water pressure is a function of y only; whence after integration of the second equation:

$$p\,(y) = C - \rho_w g\; y \tag{2.64}$$

where C is an integration constant, which is readily solved by considering that the water pressure on the water surface $y = mH$ is zero:

$$p\,(y = mH) = 0 \Rightarrow C = \rho_w g\; mH \tag{2.65}$$

The stress in the water thus reads (application of (1.12)):

$$\underline{\underline{\sigma}}^w = \rho_w g\,(y - mH) \begin{bmatrix} 1 & 0 & 0 \\ & 1 & 0 \\ \text{sym} & & 1 \end{bmatrix} \tag{2.66}$$

The surface tractions the water exerts on the surface OB' is $\overrightarrow{t}^{\,d} = \underline{\underline{\sigma}}^w \cdot \overrightarrow{n} = \underline{\underline{\sigma}}^w \cdot \overrightarrow{e}_x$:

$$\overrightarrow{t}^{\,d} = \underline{\underline{\sigma}} \cdot \overrightarrow{n} = \rho_w g\,(y - mH) \begin{bmatrix} 1 & 0 & 0 \\ & 1 & 0 \\ \text{sym} & & 1 \end{bmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \rho_w g\,(y - mH)\,\overrightarrow{e}_x \tag{2.67}$$

We can now turn to the determination of the potential energy (2.55), which reads here:

$$\begin{aligned} \mathcal{E}_{pot}\,(\overrightarrow{u}^*) &= \int_V \psi\left(\underline{\underline{\varepsilon}}^*\right) dV - \left(\int_V \overrightarrow{u}^* \cdot \rho_d \overrightarrow{g}\, dV + \int_S \overrightarrow{u}^* \cdot \overrightarrow{t}^{\,d}\, dS \right) && (2.68) \\ &= \int_{x=0}^{x=H\tan\alpha} \int_{y=0}^{y=H} \psi\left(\underline{\underline{\varepsilon}}^*\right) dxdy && (2.69) \\ &\quad + \int_{x=0}^{x=H\tan\alpha} \int_{y=0}^{y=H} \rho_d g\,(\overrightarrow{u}^* \cdot \overrightarrow{e}_y)\, dxdy && (2.70) \\ &\quad - \int_{y=0}^{y=mH} \rho_w g\,(y - mH)\,(\overrightarrow{u}^* \cdot \overrightarrow{e}_x)\; dy && (2.71) \end{aligned}$$

In some more details:
The free energy in the dam needs to be evaluated from (1.17), and requires first the determination of the strain $\underline{\underline{\varepsilon}}^*$ from (1.9). The two non-zero displacements in (2.59) read:

$$u_x^* = a\frac{y}{H};\; u_y^* = b\frac{y}{H} \tag{2.72}$$

The strains are obtained by substituting (2.72) in (1.9):

$$\underline{\underline{\varepsilon}}^* = \begin{bmatrix} \varepsilon^*_{xx} = \dfrac{\partial u^*_x}{\partial x} = 0 & \varepsilon_{xy} = \dfrac{1}{2}\left[\dfrac{\partial u_x}{\partial y} + \dfrac{\partial u_y}{\partial x}\right] = \dfrac{a}{2H} & \varepsilon_{xz} = \dfrac{1}{2}\left[\dfrac{\partial u_x}{\partial z} + \dfrac{\partial u_z}{\partial x}\right] = 0 \\ & \varepsilon^*_{yy} = \dfrac{\partial u_y}{\partial y} = \dfrac{b}{H} & \varepsilon_{yz} = \dfrac{1}{2}\left[\dfrac{\partial u_y}{\partial z} + \dfrac{\partial u_z}{\partial y}\right] = 0 \\ \text{sym} & & \varepsilon^*_{zz} = \dfrac{\partial u_z}{\partial y} = 0 \end{bmatrix} \tag{2.73}$$

The strain invariants in (1.17) read:

$$\lim tr\underline{\underline{\varepsilon}}^* = \varepsilon^*_{xx} + \varepsilon^*_{yy} + \varepsilon^*_{zz} = \frac{b}{H} \tag{2.74}$$

$$\begin{aligned} \lim tr\left(\underline{\underline{\varepsilon}}^* \cdot \underline{\underline{\varepsilon}}^*\right) &= \lim tr\left(\begin{bmatrix} 0 & \frac{a}{2H} & 0 \\ \frac{a}{2H} & \frac{b}{H} & 0 \\ 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 0 & \frac{a}{2H} & 0 \\ \frac{a}{2H} & \frac{b}{H} & 0 \\ 0 & 0 & 0 \end{bmatrix}\right) \\ &= \lim tr \begin{bmatrix} \frac{1}{4}\frac{a^2}{H^2} & \frac{1}{2}\frac{a}{H^2}b & 0 \\ \frac{1}{2}\frac{a}{H^2}b & \frac{1}{4}\frac{a^2}{H^2} + \frac{b^2}{H^2} & 0 \\ 0 & 0 & 0 \end{bmatrix} \\ &= \frac{1}{4}\frac{a^2}{H^2} + \frac{1}{4}\frac{a^2}{H^2} + \frac{b^2}{H^2} = \frac{1}{2}\frac{a^2}{H^2} + \frac{b^2}{H^2} \end{aligned} \tag{2.75}$$

Using (2.74) and (2.75) in (1.17) yields:

$$\psi = \frac{\lambda}{2}\left(\lim tr\underline{\underline{\varepsilon}}\right)^2 + \mu \lim tr\left(\underline{\underline{\varepsilon}} \cdot \underline{\underline{\varepsilon}}\right) = \frac{\lambda}{2}\left(\frac{b}{H}\right)^2 + \mu\left(\frac{1}{2}\frac{a^2}{H^2} + \frac{b^2}{H^2}\right) \tag{2.76}$$

The total free energy (2.69), therefore, which is contained in the dam reads:

$$\int_{x=0}^{x=H\tan\alpha}\int_{y=0}^{y=H} \psi\left(\underline{\underline{\varepsilon}}^*\right) dxdy = A\left(\frac{\lambda}{2}\left(\frac{b}{H}\right)^2 + \mu\left(\frac{1}{2}\frac{a^2}{H^2} + \frac{b^2}{H^2}\right)\right) \tag{2.77}$$

where $A = \frac{1}{2}H^2\tan\alpha$ is the section of the dam. This quantity can be rewritten in the form of the finite element procedure:

$$\frac{1}{2}u_i K_{ij} u_j = \frac{1}{2}(a,b)\begin{bmatrix} A\dfrac{\mu}{H^2} & 0 \\ 0 & A\dfrac{\lambda+2\mu}{H^2} \end{bmatrix}\begin{pmatrix} a \\ b \end{pmatrix} \tag{2.78}$$

The external work contributions to the potential energy (2.70) and (2.71) are developed in the form:

$$\int_{x=0}^{x=H\tan\alpha}\int_{y=0}^{y=H} \rho_d g\, u^*_y\, dxdy = A\frac{\rho_d g}{3}b \tag{2.79}$$

$$-\int_{y=0}^{y=mH} \rho_w g\,(y - mH)\, u^*_x\, dy = -\frac{\rho_w g m^3 H^2}{6}a \tag{2.80}$$

or in terms of the finite element procedure:

$$-F_i^d u_i = -(a,b) \begin{pmatrix} \dfrac{\rho_w g m^3 H^2}{6} \\ -A\dfrac{\rho_d g}{3} \end{pmatrix} \tag{2.81}$$

The potential energy (2.57) is the sum of (2.78) and (2.81):

$$\mathcal{E}_{pot}(\overrightarrow{u}^*) = \frac{1}{2}(a,b) \begin{bmatrix} A\dfrac{\mu}{H^2} & 0 \\ 0 & A\dfrac{\lambda+2\mu}{H^2} \end{bmatrix} \begin{pmatrix} a \\ b \end{pmatrix} - (a,b) \begin{pmatrix} \dfrac{\rho_w g m^3 H^2}{6} \\ -A\dfrac{\rho_d g}{3} \end{pmatrix} \tag{2.82}$$

Minimization (according to (2.58)) yields the linear system of equations:

$$\frac{\partial \mathcal{E}_{pot}(a,b)}{\partial (a,b)} = 0 \Rightarrow \begin{bmatrix} A\dfrac{\mu}{H^2} & 0 \\ 0 & A\dfrac{\lambda+2\mu}{H^2} \end{bmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \dfrac{\rho_w g m^3 H^2}{6} \\ -A\dfrac{\rho_d g}{3} \end{pmatrix} \tag{2.83}$$

The two unknowns of the problem (a,b) that minimize the potential energy are:

$$\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \dfrac{\rho_w g m^3 H^2}{3\tan\alpha\mu} \\ -\dfrac{\rho_d g H^2}{3(\lambda+2\mu)} \end{pmatrix} \tag{2.84}$$

The displacement a is the horizontal crest displacement and the displacement b is the vertical displacement.
To evaluate an order of magnitude, let us consider a concrete dam, for which the Young's modulus is roughly $E = 30$ GPa, the Poisson's ratio $\nu = 0.2$, so that the Lamé constants (1.18) are $\lambda = 8.3$ GPa and $\mu = 12.5$ GPa. The water density is $\rho_w = 1,000$ kg/m^3, and the concrete density is $\rho_d = 2,500$ kg/m^3. The crest and base widths of the dam displayed in figure 2.4 is 24 m and 140 m, so that $\tan\alpha = (140-24)/H$. A lower bound estimate of the crest displacement is $u_x^*(y=H) = 2.3 \times 10^{-9}(mH)^3$; for instance, for a 100 m high dam, which is fully filled ($m = 1$), a lower bound of the crest displacement is on the order of 3 mm (quite negligible!).

Finite Element Solution

Congratulations! We just approximated the gravity dam by one single 2-D triangle element. As we refine the discretization by adding more triangle elements we come closer to the actual potential energy in the gravity dam. This is readily achieved by any commercially available finite element program. This is to be shown for $m = 1$ (end of water filling), and by setting the self weight of the dam to zero. The obtained crest displacement is normalized by the solution of the one-triangle element solution (2.84).

5.3 Concluding Remarks

The finite element method is no-doubt an extremely powerful engineering method, because of its flexibility to be adapted to many engineering problems. The important issues to consider are:

1. The starting point for the development of any FE-procedure is the physics of the continuum problem. This includes: (1) choice of a system and the boundary conditions; (2) the filed equations derived from conservation laws, and (3) the constitutive information.

2. The FE-method explores the convexity of the constitutive law. It is on the basis of the convexity that one can translate an 'exact' problem into an 'approximated' problem. Because of the convexity the approximation should always converge to the exact solution.

Chapter 6

Discrete to Continuum Modeling

These notes give a few examples illustrating how continuum models can be derived from special limits of discrete models. Only the simplest cases are considered, illustrating some of the most basic ideas. These techniques are useful because continuum models are often much easier to deal with than discrete models with very many variables, both conceptually and computationally.

6.1 Introduction.

Continuum approximations are useful in describing discrete systems with a large number of degrees of freedom. In general, a continuum approximation will not describe all possible solutions of the discrete system, but some special class that will depend on the approximations and assumptions made in deriving the continuum model. Whether or not the approximation is useful in describing a particular situation, will depend on the appropriate approximations being made. The most successful models arise in situations where most solutions of the discrete model evolve rapidly in time towards configurations where the assumptions behind the continuum model apply.

The basic step in obtaining a continuum model from a discrete system, is to identify some basic configuration (solution of the discrete model) that can be described by a few parameters. Then one assumes that the full solution of the system can be described, near every point in space and at every time, by this configuration — for some value of the parameters. The parameters are then assumed to vary in space and time, but on scales (macro-scales) that are much larger than the ones associated with the basic configuration (micro-scales). Then one attempts to derive equations describing the evolution of these parameters in the macro-scales, thus averaging out of the problem the micro-scales. There is a close connection between this approach, and the "quasi-equilibrium" approximations that are often invoked to "close" continuum sets of equations derived using conservation laws.

For example, when deriving the equations for Gas Dynamics in Statistical Mechanics, it is assumed that the local particle interactions rapidly exchange energy and momentum between the molecules — so that the local probability distributions for velocities take a standard form (equivalent to local thermodynamic equilibrium). What exactly makes these assumptions work (in terms of properties of the governing, micro-scale, equations) is rather poorly understood. But that they work rather well cannot be denied. In these notes we will consider examples that are rather simpler than these ones, however, where the "local configurations" tend to be rather trivial.

6.2 Wave Equations from Mass-Spring Systems.

Longitudinal Motion.

Consider an array of bodies/particles, connected by springs, and restricted[1] to move on a straight line. Let the **positions of the bodies** be given by $x_n = x_n(t)$, with $n = 0, \pm 1, \pm 2, \ldots$, and let M_n be the **mass of the n^{th} particle.** Furthermore, let the **force law for the spring** between particles n and $n+1$ be given by: force $= f_{n+\frac{1}{2}}(\Delta x)$, where Δx is the distance between the particles, and $f_{n+\frac{1}{2}}$ is **positive when the spring is under tension.**[2]

If there are no other forces involved (e.g. no friction), the governing equations for the system are:

$$M_n \frac{d^2}{dt^2} x_n = f_{n+\frac{1}{2}}(x_{n+1} - x_n) - f_{n-\frac{1}{2}}(x_n - x_{n-1}) \,, \tag{2.1}$$

for $n = 0, \pm 1, \pm 2, \ldots$ The simplest solution for this system of equations is equilibrium. In this case all the accelerations vanish, so that the particle positions are given by the series of algebraic equations

$$0 = f_{n+\frac{1}{2}}(x_{n+1} - x_n) - f_{n-\frac{1}{2}}(x_n - x_{n-1}) \,. \tag{2.2}$$

This is the basic configuration (solution) that we will use in obtaining a continuum approximation. Note that this is a one parameter family: if the forces are monotone functions of the displacements Δx, then once any one of them is given, the others follow from (2.2).

Before proceeding any further, it is a good idea to **non-dimensionalize the equations.** We will **assume** that:

A. All the springs are roughly similar, so that we can talk of a typical spring force f, and a typical spring length L. Thus we can write

$$f_{n+\frac{1}{2}}(\Delta x) = f\, F_{n+\frac{1}{2}}\left(\frac{\Delta x}{L}\right) \,, \tag{2.3}$$

where $F_{n+\frac{1}{2}}$ is a non-dimensional mathematical function, of $O(1)$ size, and with $O(1)$ derivatives. A further assumption is that $F_{n+\frac{1}{2}}$ changes slowly with n, so that two nearby springs are nearly equal. Mathematically, this is specified by stating that:

$$F_{n+\frac{1}{2}}(\eta) = F(\epsilon(n+1/2), \eta) \,, \tag{2.4}$$

where $0 < \epsilon \ll 1$, and F is a "nice" (mathematical) function of its two variables.

B. All the **particles have roughly the same mass m, and their masses change slowly with n,** so that we can write:

$$M_n = m\, M(\epsilon\, n) \,, \tag{2.5}$$

where M is a nice mathematical function, with $O(1)$ size, and with $O(1)$ derivatives.

---

[1] By some device: say the bodies are sliding inside a hollow tube.

[2] If the spring obeys Hooke's law, then $f_{n+\frac{1}{2}}(\Delta x) = k_{n+\frac{1}{2}}\left(\Delta x - L_{n+\frac{1}{2}}\right)$, where $k_{n+\frac{1}{2}} > 0$ and $L_{n+\frac{1}{2}} > 0$ are the spring constant and equilibrium length, respectively.

Remark 6.2.1 *Why do we need these assumptions? This has to do with the questions of validity, discussed in the introduction. Suppose that these hypothesis are violated, with the masses and springs jumping wildly in characteristics. Then the basic configuration described by (2.2) will still be a solution. However, as soon as there is any significant motion, neighboring parts of the chain will respond very differently, and the solution will move away from the local equilibrium implied by (2.2). There is no known method to, generically, deal with these sort of problems — which turn out to be very important: see remark 6.2.2.*

From the assumptions in **A** and **B** above, we see that:

$$\text{Changes in the mass-spring system occur over length scales} \quad \boxed{\ell = L/\epsilon.} \tag{2.6}$$

Using this scale to non-dimensionalize space, namely: $\boxed{x_n = \ell\, X_n,}$ and a yet to be specified time scale τ to non-dimensionalize time, namely: $\boxed{t = \tau\, T,}$ the equations become:

$$M(\epsilon n)\,\frac{d^2}{dT^2}X_n = \frac{\epsilon\, f\, \tau^2}{m\, L}\,\left(F_{n+\frac{1}{2}}\left(\frac{X_{n+1}-X_n}{\epsilon}\right) - F_{n-\frac{1}{2}}\left(\frac{X_n - X_{n-1}}{\epsilon}\right)\right)\,. \tag{2.7}$$

A and **B** above also imply that, for the solution in (2.2), the inter-particle distance $x_{n+1} - x_n$ varies slowly — an $O(\epsilon)$ fractional amount per step in n. Thus we propose solutions for (2.7) of the form:

$$X_n(t) = X(s_n,\, T)\,, \quad \text{where} \quad s_n = n\,\epsilon\,, \tag{2.8}$$

and $X = X(s,T)$ is some smooth function of its arguments.

Substituting (2.8) into (2.7), and using (2.4) and (2.5), we obtain

$$M(s)\,\frac{\partial^2}{\partial T^2}X = \frac{\epsilon^2\, f\, \tau^2}{m\, L}\,\left(\frac{\partial}{\partial s}F\left(s,\, \frac{\partial}{\partial s}X\right) + O(\epsilon^2)\right)\,. \tag{2.9}$$

Here we have used that:

$$\frac{X_{n+1}-X_n}{\epsilon} = \frac{\partial}{\partial s}X(s_n + \frac{1}{2}\epsilon, t) + O(\epsilon^2) \quad \text{and} \quad \frac{X_n - X_{n-1}}{\epsilon} = \frac{\partial}{\partial s}X(s_n - \frac{1}{2}\epsilon, t) + O(\epsilon^2)\,,$$

with a similar formula applying to the difference $F_{n+\frac{1}{2}} - F_{n-\frac{1}{2}}$.

Equation (2.9) suggests that we should take

$$\tau = \sqrt{\frac{m\, L}{\epsilon^2\, f}}\,, \tag{2.10}$$

for the un-specified time scale in (2.7). Then equation (2.9) leads to the **continuum limit approximations** (valid for $0 < \epsilon \ll 1$)

$$M(s)\,\frac{\partial^2}{\partial T^2}X = \frac{\partial}{\partial s}F\left(s,\, \frac{\partial}{\partial s}X\right)\,. \tag{2.11}$$

The mass-spring system introduced in equation (2.1) can be thought of as a simple model for an elastic rod under (only) longitudinal forces. Then we see that (2.11) is a model **(nonlinear wave) equation for the longitudinal vibrations of an elastic rod,** with s a lagrangian coordinate for the points in the rod, $M = M(s)$ the mass density along the rod, and X giving the position of the point s as a function of time, and F a function characterizing the elastic response of the rod. Of course, in practice F must be obtained from laboratory measurements.

Remark 6.2.2 *The way in which the equations for nonlinear elasticity can be derived for a crystalline solid is not too different*[3] *from the derivation of the wave equation (2.11) for longitudinal vibrations. Then a very important question arises (see first paragraph in section 6.1): What important behaviors are missed due to the assumptions in the derivation? How can they be modeled? In particular, what happens if there are "defects" in the crystal structure (see remark 6.2.1)? These are all very important, and open, problems of current research interest.*

Example 6.2.1 Uniform Rod.
If all the springs and all the particles are equal, then we can take $M \equiv 1$ and F is independent of s. Furthermore, if we take L to be the (common) equilibrium length of the springs, we then have

$$\frac{\partial^2}{\partial T^2} X = \frac{\partial}{\partial s} F\left(\frac{\partial}{\partial s} X\right) = c^2 \left(\frac{\partial}{\partial s} X\right) \frac{\partial^2}{\partial s^2} X\,, \tag{2.12}$$

where $c^2 = c^2(\eta) = dF/d\eta(\eta) > 0$, and $F(1) = 0$ (equilibrium length). The unperturbed "rod" corresponds to $X \equiv s$, while $X \equiv \alpha\, s$ corresponds to the rod under uniform tension ($\alpha > 1$), or compression ($\alpha < 1$). Also, note that c is a (non-dimensional) speed — the speed at which elastic disturbances along the rod propagate: i.e. the **sound speed.**

Example 6.2.2 Small Disturbances.
Consider a uniform rod in a situation where the departures from uniform equilibrium are small. That is $\partial X/\partial s \approx \alpha$, where α is a constant. Then equation (2.12) can be approximated by the **linear wave equation**

$$X_{TT} = c^2\, X_{ss}\,, \tag{2.13}$$

where $c = c(\alpha)$ is a constant. The general solution to this equation has the form

$$X = g(s - c\,T) + h(s + c\,T)\,, \tag{2.14}$$

where g and h are arbitrary functions. This solution clearly shows that c is the wave propagation velocity.

Remark 6.2.3 Fast vibrations.
The vibration frequency for a typical mass m, attached to a typical spring in the chain, is:

$$\omega = \sqrt{\frac{f}{m\,L}} = \frac{1}{\epsilon\tau}\,. \tag{2.15}$$

This corresponds to a time scale much shorter than the one involved in the solution in (2.8–2.11). What role do the motions in these scales play in the behavior of the solutions of (2.1), *under the assumptions made earlier in* **A** *and* **B***?*

[3] At least qualitatively, though it is technically far more challenging.

For real crystal lattices, which are definitely not one dimensional *(as the one in (2.1)) these fast time scales correspond to thermal energy (energy stored in the local vibrations of the atoms, relative to their equilibrium positions). It is believed that the nonlinearities in the lattice act so as to randomize these vibrations, so that the energy they contain propagates as heat (diffuses). In one dimension, however, this does not generally happen, with the vibrations remaining coherent enough to propagate with a strong wave component. The actual processes involved are very poorly understood, and the statements just made result, mainly, from numerical experiments with nonlinear lattices.*

Just to be a bit more precise: consider the situation where all the masses are equal — $\boxed{M_n = m}$ *for all n, and all the springs are equal and satisfy Hooke's law (linear elasticity):*

$$f_{n+\frac{1}{2}}(\Delta x) = k(\Delta x - L) = f\left(\frac{\Delta x}{L} - 1\right), \tag{2.16}$$

where k is the spring constant, L is the equilibrium length, and $\boxed{f = k\,L.}$ *Then equation (2.1) takes the form*

$$\frac{d^2}{dt^2}x_n = \omega^2\left(x_{n+1} - 2x_n + x_{n-1}\right), \tag{2.17}$$

where ω is as in (2.15). Because this system is linear, we can write its general solution as a linear superposition of eigenmodes, which are solutions of the form[4]

$$x_n = \exp(i\,\kappa\,n - i\,\sigma\,t), \quad \textit{where } \sigma = \pm 2\,\omega\,\sin\left(\frac{\kappa}{2}\right) \textit{ and } -\infty < \kappa < \infty \textit{ is a constant.} \tag{2.18}$$

These must be added to an equilibrium solution $x_n = \alpha\,L\,n = s_n$, where $\alpha > 0$ is a constant.

Relative to the mean position s_n along the lattice, each solution in (2.18) can be written as

$$x_n = \exp(i\,\frac{\kappa}{\alpha L}\,s_n - i\,\sigma\,t).$$

Thus we see that it represents a wave of wavelength $\lambda = 2\pi\alpha L/\kappa$, and speed

$$c_w = \frac{\alpha L\sigma}{\kappa} = \pm\frac{2\alpha L\omega}{\kappa}\sin\left(\frac{\kappa}{2}\right) = \frac{2c}{\kappa}\sin\left(\frac{\kappa}{2}\right) \tag{2.19}$$

propagating along the lattice — where $c = \alpha L\omega$ is a speed. Note that the speed of propagation is a function of the wave-length — this phenomenon is know by the name of **dispersion**. *We also note that the maximum frequency these eigenmodes can have is $\sigma = 2\omega$, and corresponds to wavelengths of the order of the lattice separation.*[5]

In the case of equations (2.16 – 2.17) there is no intrinsic ϵ in the equations: it must arise from the initial conditions. That is to say: assume that the wavelength ℓ with which the lattice is excited is much larger than the lattice equilibrium separation L, i.e. $\ell \gg L$, with $\epsilon = L/\ell$. This corresponds to solutions (2.18) with κ small. In this **long wave limit** *we see that (2.19) implies that the solutions have the same wave speed $c_w = \pm c$. This corresponds to the situation in (2.13 – 2.14).*

It is clear that, in the linear lattice situation described above, we cannot dismiss the fast vibration excitations (with frequencies of the order of ω) as constituting some sort of energy "bath" to be

[4] Check that these are solutions.

[5] The reason for the 2 relative to (2.15) is that the masses are coupled, and not attached to a single spring.

interpreted as heat. The energy in these vibrations propagates as waves through the media, with speeds which are of the same order of magnitude as the sound waves equation (2.13) describes. Before the advent of computers it was believed that nonlinearity would destroy the coherence of these fast vibrations. Numerical experiments, however, have shown that this is not (generally) true for one dimensional lattices,[6] *though it seems to be true in higher dimensions. Exactly why, and how, this happens is a subject of some current interest.*

Transversal Motion.

We consider now a slightly different situation, in which the masses are allowed to move only in the direction perpendicular to the x axis. To be precise: consider a sequence of masses M_n in the plane, whose x coordinates are given by $x_n = n\,L$. Each mass is restricted to move only in the orthogonal coordinate direction, with $y_n = y_n(t)$ giving its y position. The masses are connected by springs, with $f_{n+\frac{1}{2}}(\Delta r_{n+\frac{1}{2}})$ the force law, where $\Delta r_{n+\frac{1}{2}} = \sqrt{L^2 + (y_{n+1} - y_n)^2}$ is the distance between masses. Assuming that there are no other forces involved, the governing equations for the system are:

$$M_n \frac{d^2}{dt^2} y_n = \frac{y_{n+1} - y_n}{\Delta r_{n+\frac{1}{2}}} f_{n+\frac{1}{2}}(\Delta r_{n+\frac{1}{2}}) - \frac{y_n - y_{n-1}}{\Delta r_{n-\frac{1}{2}}} f_{n-\frac{1}{2}}(\Delta r_{n-\frac{1}{2}})\,, \tag{2.20}$$

for $n = 0,\ \pm 1,\ \pm 2, \ldots$ (you should convince yourself that this is the case).

The simplest solution for this system of equations is equilibrium, with all the masses lined up horizontally $\boxed{y_{n+1} = y_n,}$ so that **all the accelerations vanish.** Again, one can use this (one parameter) family of solutions to obtain a continuum approximation for the system in (2.20) — under the same assumptions earlier in **A** and **B**.

Remark 6.2.4 Stability of the Equilibrium Solutions.
It should be intuitively obvious that the equilibrium solutions described above will be stable only if the equilibrium lengths of the springs $\mathcal{L}_{n+\frac{1}{2}}$ are smaller than the horizontal separation L between the masses, namely: $\mathcal{L}_{n+\frac{1}{2}} < L$. *This so that none of the springs is under compression in the solution, since any mass in a situation where its springs are under compression will easily "pop" out of alignment with the others — see example 6.2.3.*

Introduce now the non-dimensional variables $Y = \epsilon\, y/L$, $X = \epsilon\, x/L$ (note that, since $x_n = nL$, in fact X plays here the same role that s played in the prior derivation[7]), and $T = t/\tau$, where τ is as in (2.10). Then the **continuum limit for the equations in (2.20)** is given by

$$M(X)\,\frac{\partial^2 Y}{\partial T^2} = \frac{\partial}{\partial X}\left(\frac{F(X,\mathcal{S})}{\mathcal{S}}\,\frac{\partial Y}{\partial X}\right) \tag{2.21}$$

where $Y = Y(X,T)$ and

$$\mathcal{S} = \sqrt{1 + \left(\frac{\partial Y}{\partial X}\right)^2}\,.$$

[6] The first observation of this general phenomena was reported by E. Fermi, J. Pasta and S. Ulam, in 1955: *Studies of Non Linear Problems*, Los Alamos Report LA-1940 (1955), pp. 978-988 in Collected Papers of Enrico Fermi. II, The University of Chicago Press, Chicago, (1965).

[7] The coordinate s is simply a label for the masses. Since in this case the masses do not move horizontally, X can be used as the label.

The derivation of this equation is left as an exercise to the reader.

The mass-spring system introduced in (2.20) can be thought of as a simple model for an elastic string restricted to move in the transversal direction only. Then we see that (2.21) is a model **(nonlinear wave) equation for the transversal vibrations of a string**, where X is the longitudinal coordinate along the string position, Y is the transversal coordinate, $M = M(X)$ is the mass density along the string, and $F = F(X, \mathcal{S})$ describes the elastic properties of the string.[8] In the non-dimensional coordinates, the (local) equilibrium length for the string is given by $e_\ell = \mathcal{L}/L$. That is, the elastic forces vanish for this length:

$$F(X, e_\ell(X)) \equiv 0\,, \quad \text{where} \quad e_\ell < 1 \quad \text{(for stability, see remark 6.2.4).} \tag{2.22}$$

We also assume that $\boxed{\dfrac{\partial}{\partial \mathcal{S}} F(X, \mathcal{S}) > 0.}$

Example 6.2.3 Uniform String with Small Disturbances.
Consider now a uniform string (neither M, nor F, depend on X) in a situation where the departures from equilibrium are small ($\partial Y/\partial X$ is small).

For a uniform string we can assume $M \equiv 1$, and F is independent of X. Thus equation (2.21) reduces to

$$\frac{\partial^2 Y}{\partial T^2} = \frac{\partial}{\partial X}\left(\frac{F(\mathcal{S})}{\mathcal{S}}\,\frac{\partial Y}{\partial X}\right). \tag{2.23}$$

Next, for small disturbances we have $\mathcal{S} \approx 1$, and (2.23) can be approximated by the **linear wave equation**

$$Y_{TT} = c^2\, Y_{XX}\,, \tag{2.24}$$

where $c^2 = F(1)$ is a constant (see equations (2.13 – 2.14).

Notice how the stability condition $e_\ell < 1$ in (2.22) guarantees that $c^2 > 0$ in (2.23). If this were not the case, instead of the linear wave equation, the linearized equation would have been of the form

$$Y_{TT} + d^2\, Y_{XX} = 0\,, \tag{2.25}$$

with $d > 0$. This is **Laplace Equation**, *which is* **ill-posed as an evolution in time problem.** *To see this, it is enough to notice that (2.25) has the following solutions:*

$$Y = e^{d\,|k|\,t}\,\sin(kX)\,, \quad \textit{for any } -\infty < k < \infty\,. \tag{2.26}$$

These solutions grow arbitrarily fast in time, the fastest the shortest the wave-length ($|k|$ larger). This is just the mathematical form of the obvious physical fact that a straight string (with no bending strength) is not a very stable object when under compression.

General Motion: Strings and Rods.

[8] Notice that $\mathcal{S}$ is the local stretching of the string, due to its inclination relative to the horizontal position (actual length divided by horizontal length).

If no restrictions to longitudinal (as in (2.1)) or transversal (as in (2.20)) motion are imposed on the mass-spring chain, then (in the continuum limit) general equations including both longitudinal and transversal modes of vibration for a string are obtained. Since strings have no bending strength, these equations will be well behaved only as long as the string is under tension everywhere.

Bending strength is easily incorporated into the mass-spring chain model. Basically, what we need to do is to incorporate, at the location of each mass point, a bending spring. These springs apply a torque when their ends are bent, and will exert a force when-ever the chain is not straight. The continuum limit of a model like this will be equations describing the vibrations of a rod.

We will not develop these model equations here.

6.3 Torsion Coupled Pendulums: Sine-Gordon Equation.

Consider an horizontal axle A, of total length ℓ, suspended at its ends by "frictionless" bearings. Along this axle, at equally spaced intervals, there are N equal pendulums. Each pendulum consists of a rigid rod, attached perpendicularly to the axle, with a mass at the end. When at rest, all the pendulums point down the vertical. We now make the following assumptions and approximations:

- **1.** Each pendulum has a mass $\dfrac{M}{N}$. The distance from its center of mass to the axle center is L.
- **2.** The axle A is free to rotate, and we can ignore any frictional forces (i.e.: they are small). In fact, the only forces that we will consider are gravity, and the torsional forces induced on the axle when the pendulums are not all aligned.
- **3.** Any deformations to the axle and rod shapes are small enough that we can ignore them. Thus the axle and rod are assumed straight at all times.
- **4.** The mass of the axle is small compared to M, so we ignore it (this assumption is not strictly needed, but we make it to keep matters simple).

Our aim is to produce a continuum approximation for this system, as $N \to \infty$, with everything else fixed.

Each one of the **pendulums can be characterized by the angle $\theta_n = \theta_n(t)$ that its suspending rod makes with the vertical direction.** Each pendulum is then subject to **three forces:**

(a) Gravity, for which only the component perpendicular to the pendulum rod is considered.[9]
(b) Axle torsional force due to the twist $\theta_{n+1} - \theta_n$. This couples each pendulum to the next one.
(c) Axle torsional force due to the twist $\theta_n - \theta_{n-1}$. This couples each pendulum to the prior one.

We will assume that the amount of twist per unit length in the axle is small, so that Hooke's law applies.

Remark 6.3.1 Hooke's Law for Torsional Forces.
In the Hooke's law regime, for a given fixed bar, the torque generated is directly proportional to the angle of twist, and inversely proportional to the distance over which the twist occurs.
To be specific: in the problem here, imagine that a section of length $\Delta\ell$ of the axle has been twisted by an amount (angle) Ψ. Then, if T is the torque generated by this twist, one can write

$$T = \frac{\kappa\,\Psi}{\Delta\ell}\,, \tag{3.1}$$

[9] The component along the rod is balanced by the rod itself, which we approximate as being rigid.

where κ **is a constant** *that depends on the axle material and the area of its cross-section* — assume that the axle is an homogeneous cylinder. **The dimensions of** κ **are given by:**

$$[\kappa] = \frac{mass \times length^3}{time^2 \times angle} = \frac{force \times area}{angle}\,. \tag{3.2}$$

This torque then translates onto a tangential force of magnitude $F = T/L$*, on a mass attached to the axle at a distance* L*. The sign of the force is such that it opposes the twist.*

Let us now go back to our problem, and write the equations of motion for the N pendulums. We will assume that:

- The horizontal separation between pendulums is $\dfrac{\ell}{N+1}$.
- The first and last pendulum are at a distance $\dfrac{\ell}{2(N+1)}$ from the respective ends of the axle.

The tangential force (perpendicular to the pendulum rod) due to gravity on each of the masses is

$$F_g = -\frac{1}{N}\, Mg \sin\theta_n\,, \quad \text{where} \quad n = 1, \ldots, N\,. \tag{3.3}$$

For any two successive masses, there is also a torque whenever $\theta_n \neq \theta_{n+1}$. This is generated by the twist in the axle, of magnitude $\theta_{n+1} - \theta_n$, over the segment of length $\ell/(N+1)$ connecting the two rods. Thus each of the masses experiences a force (equal in magnitude and opposite in sign)

$$F_T = \pm\,(N+1)\frac{\kappa}{\ell L}\,(\theta_{n+1} - \theta_n)\,, \tag{3.4}$$

where the signs are such that the forces tend to make $\theta_n = \theta_{n+1}$. Putting all this together, we obtain the following set of equations for the angles:

$$\frac{1}{N}\, ML\, \frac{d^2\theta_1}{dt^2} = -\frac{1}{N}\, Mg \sin\theta_1 + \frac{(N+1)\,\kappa}{\ell L}\,(\theta_2 - \theta_1)\,, \tag{3.5}$$

$$\begin{aligned} \frac{1}{N}\, ML\, \frac{d^2\theta_n}{dt^2} &= -\frac{1}{N}\, Mg \sin\theta_n \\ &\quad + \frac{(N+1)\,\kappa}{\ell L}\,(\theta_{n+1} - \theta_n) - \frac{(N+1)\,\kappa}{\ell L}\,(\theta_n - \theta_{n-1})\,, \\ &\quad \text{for } n = 2, \ldots, N-1, \text{ and} \end{aligned} \tag{3.6}$$

$$\frac{1}{N}\, ML\, \frac{d^2\theta_N}{dt^2} = -\frac{1}{N}\, Mg \sin\theta_N - \frac{(N+1)\,\kappa}{\ell L}(\theta_N - \theta_{N-1})\,. \tag{3.7}$$

These are the **equations for** N **torsion coupled equal pendulums.**

Remark 6.3.2 *To check that the signs for the torsion forces selected in these equations are correct, take the difference between the* n^{th} *and* $(n+1)^{th}$ *equation. Then you should see that the torsion force (due to the portion of the axle connecting the* n^{th} *and* $(n+1)^{th}$ *pendulums) is acting so as to make the angles equal.*

Remark 6.3.3 *Note that the equations for the first and last angle are different, because the first and last pendulum experience a torsion force from only one side.* **How would you modify these equations to account for having one (or both) ends of the axle fixed?**

Continuum Limit.

Now we consider the continuum limit, in which we let $N \to \infty$ and assume that the n^{th} angle can be written in the form:

$$\theta_n(t) = \theta(x_n,\, t)\,, \tag{3.8}$$

where $\theta = \theta(x,t)$ is a "nice" function (with derivatives) and $x_n = \dfrac{n+\frac{1}{2}}{N+1}\,\ell$ is the position of the pendulum along the axle. In particular, note that:

$$\Delta x = x_{n+1} - x_n = \frac{\ell}{N+1}\,. \tag{3.9}$$

Take equation (3.6), and multiply it by N/ℓ. Then we obtain

$$\rho\, L\, \frac{d^2\theta_n}{dt^2} = -\rho\, g \sin\theta_n + \frac{N(N+1)\kappa}{\ell^2 L}\,(\theta_{n+1} - 2\theta_n + \theta_{n-1})\,,$$

where $\boxed{\rho = M/\ell}$ is the **mass density per unit length** in the $N \to \infty$ limit. Using equation (3.9), this can be written in the form:

$$\rho\, L\, \frac{d^2\theta_n}{dt^2} = -\rho\, g \sin\theta_n + \frac{N}{(N+1)}\,\frac{\kappa}{L}\,\frac{\theta_{n+1} - 2\theta_n + \theta_{n-1}}{(\Delta x)^2}\,. \tag{3.10}$$

From equation (3.8) we see that — in the limit $N \to \infty$ (where $\Delta \to 0$) — we have:

$$\frac{\theta_{n+1} - 2\theta_n + \theta_{n-1}}{(\Delta x)^2} \to \frac{\partial^2\theta}{\partial x^2}(x_n,\, t)\,.$$

Thus, finally, we obtain (for the continuum limit) the nonlinear wave equation (the **"Sine–Gordon" equation**):

$$\theta_{tt} - c^2\theta_{xx} = -\omega^2 \sin\theta\,, \tag{3.11}$$

where $\boxed{\omega = \sqrt{\dfrac{g}{L}}}$ is the pendulum angular frequency, and $\boxed{c = \sqrt{\dfrac{\kappa}{\rho L^2}}}$ is a wave propagation speed **(check that the dimensions are correct).**

Remark 6.3.4 Boundary Conditions.
What happens with the first (3.5) and last (3.7) equations in the limit $N \to \infty$?

As above, multiply (3.5) by $1/\ell$. Then the equation becomes:

$$\frac{\rho\, L}{N}\,\frac{d^2\theta_1}{dt^2} = -\frac{\rho\, g}{N}\,\sin\theta_1 + \frac{(N+1)\kappa}{\ell^2 L}\,(\theta_2 - \theta_1) = -\frac{\rho\, g}{N}\,\sin\theta_1 + \frac{\kappa}{\ell L}\,\frac{\theta_2 - \theta_1}{\Delta x}\,.$$

Thus, as $N \to \infty$ one obtains

$$\boxed{\theta_x(0,t) = 0.}$$

This is just the statement that there are no torsion forces at the $x = 0$ end (since the axle is free to rotate there). Similarly, one obtains:

$$\boxed{\theta_x(\ell,t) = 0,}$$

at the other end of the axle. **How would these boundary conditions be modified if the axle where fixed at one (or both) ends?**

Kinks and Breathers for the Sine Gordon Equation.

Equation (3.11), whose non-dimensional form is

$$\theta_{tt} - \theta_{xx} = -\sin\theta\,, \tag{3.12}$$

has a rather interesting history. Its first appearance is not in the context of a physical context at all, but in the study of the geometry of surfaces with constant negative Gaussian curvature. Physical problems for which it has been used include: Josephson junction transmission lines, dislocation in crystals, propagation in ferromagnetic materials of waves carrying rotations in the magnetization direction, etc.[10] Mathematically, it is a very interesting because **it is one of the few physically important nonlinear partial differential equations that can be solved explicitly** (by a technique known as **Inverse Scattering**, which we will not describe here).

An important consequence of equation (3.12) exact solvability, is that it possesses **particle-like solutions, known as kinks, anti-kinks, and breathers.** These are localized traveling disturbances, which preserve their identity when they interact. In fact, the only effect of an interaction is a phase shift in the particle positions after the interaction: effectively, the "particles" approach each other, stay together briefly while they interact (this causes the "phase shift") and then depart, preserving their identities and original velocities. This can all be shown analytically, but here we will only illustrate the process, using some computational examples.

The first step is to present **analytical expressions for the various particle-like solutions** of equation (3.12). These turn out to be relatively simple to write.

Example 6.3.1 Kinks and Anti-Kinks.
Equation (3.12) has some interesting solutions, that correspond to giving the pendulums a full 2π twist (e.g.: take one end pendulum, and give it a full 2π rotation). This generates a 2π twist wave that propagates along the pendulum chain. These waves are known as kinks or anti-kinks (depending on the sign of the rotation), and can be written explicitly. In fact, they are steady wave solutions,[11] *for which the equation reduces to an O.D.E., which can be explicitly solved.*

[10] For reviews see:
A. C. Scott, 1970, *Active and Nonlinear Wave Propagation in Electronics*, Wiley Interscience, New York (page 250). Barone, A. F. Esposito, C. J. Magee, and A. C. Scott, 1971, *Theory and Applications of the Sine Gordon Equation*, Rivista del Nuovo Cimento **vol. 1**, pp. 227–267.

[11] Solutions of the form $\theta = \theta(x - c\,t)$, where c is a constant: the speed of propagation.

Let $\boxed{-1 < c < 1}$ *be a constant (kink, or anti-kink speed), and let* $\boxed{z = (x - c\,t - x_0)}$ *be a moving coordinate, where the solution is steady — the "twist" will be centered at* $x = c\,t + x_0$*, where* x_0 *is the position at time* $t = 0$*. Then the* **kink** *solution is given by*

$$\theta = 2\,\arccos\left(\frac{e^{2\,z/\beta} - 1}{e^{2\,z/\beta} + 1}\right) = 4\,\arctan\left(\exp\left(-\frac{z}{\beta}\right)\right)\,, \tag{3.13}$$

where $\boxed{\beta = \sqrt{1 - c^2}}$ *is the* kink width*. This solution represents a propagating clock-wise* 2π *rotation, from* $\theta = 2\,m\,\pi$ *as* $x \to -\infty$ *(where* m *is an integer) to* $\theta = 2\,(m-1)\,\pi$ *as* $x \to \infty$*, with most of the rotation concentrated in a region of width* $O(\beta)$ *near* $x = c\,t + x_0$*. The parameter* c *is determined (for example) by how fast the initial twist is introduced when the kink is generated.*

We note now that:

- *From (3.13) it follows that* $\theta_t = -c\,\theta_x = \dfrac{2c}{\beta}\sin\left(\dfrac{\theta}{2}\right)$*. Using this, it is easy to show that (3.13) is a solution of equation (3.12).*
- *The Sine-Gordon equation is the simplest of a "class" of models proposed for nuclear interactions. In this interpretation, the kinks are nuclear particles. Since (in the non-dimensional version (3.12)) the speed of light is* 1*, the restriction* $-1 < c < 1$ *is the relativistic restriction, and the factor* β *incorporates the usual relativistic contraction.*

The **anti-kink** *solution follows by replacing* $x \to -x$ *and* $t \to -t$ *in (3.13). It corresponds to a propagating counter-clock-wise* 2π *rotation, and it is given by*

$$\theta = 2\,\arccos\left(\frac{1 - e^{2\,z/\beta}}{1 + e^{2\,z/\beta}}\right) = 4\,\arctan\left(\exp\left(\frac{z}{\beta}\right)\right)\,. \tag{3.14}$$

The kinks and anti-kinks are very non-linear solutions. Thus, it is of some interest to study how they interact with each other. Because they are very localized solutions (non-trivial only in a small region), when their centers are far enough they can be added. Thus, numerically it is rather easy to study their interactions, by setting up initial conditions that correspond to kinks and anti-kinks far enough that they do not initially interact. Then they are followed until they collide. In the lectures the results of numerical experiments of this type will be shown (the numerical method used in the experiments is is a "pseudo-spectral" method).

Example 6.3.2 Breathers.
A different kind of interesting solution is provided by the "breathers" — which we handle next. A **breather** *is a* **wave-package** *kind of solution (an oscillatory wave, with an envelope that limits the wave to reside in a bounded region of space. These solutions vanish (exponentially) as* $x \to \pm\infty$*. This last property allows for easy numerical simulations of interactions of breathers (and kinks). One can setup initial conditions corresponding to the interaction of as many kinks and/or breathers as one may wish (limited only be the numerical resolution of the computation), simply by separating them in space.*

A breather solution is characterized by two arbitrary constants $-1 < d, V < 1$. Then define

$$\left.\begin{array}{rcl} A &=& d/\sqrt{1-d^2}\,, \\ B &=& 1/\sqrt{1-V^2}\,, \\ C &=& \sqrt{1-d^2}\,, \\ \\ p &=& C\,B\,(V\,x - t + t_0)\,, \\ q &=& d\,B\,(x - V\,t - x_0)\,, \\ \\ Q &=& A\,\sin(p)/\cosh(q)\,, \end{array}\right\} \tag{3.15}$$

where x_0 and t_0 are constants, centering the envelope and the phase, respectively. Notice that the partial derivatives of Q (with respect to p and q) are given by

$$Q_p = A\,\cos(p)/\cosh(q) \quad \textit{and} \quad Q_q = -Q\,\tanh(q)\,. \tag{3.16}$$

The breather solution (and its time derivative) is then given by:

$$\left.\begin{array}{rcl} \theta &=& 4\,\arctan(Q)\,, \\ \\ \theta_t &=& -4\,(1+Q^2)\,(C\,B\,Q_p + d\,B\,V\,Q_q)\,. \end{array}\right\} \tag{3.17}$$

The breather solution is a wave-package type of solution, with the phase controlled by p, and the envelope (causing the exponential vanishing of the solution) by q). The wave-package details are given by:

$$\left.\begin{array}{rcl} \text{speed} \ldots\ldots\ldots\ldots c_p &=& 1/V, \\ \text{period} \ldots\ldots\ldots\ldots T_p &=& 2\,\pi/(B\,C)\,, \\ \text{wave-length} \ldots\ldots\ldots \lambda_p &=& 2\,\pi/(B\,C\,V)\,, \end{array}\right\} \quad \textbf{Phase.} \tag{3.18}$$

$$\left.\begin{array}{rcl} \text{speed} \ldots\ldots\ldots\ldots c_e &=& V, \\ \text{width} \ldots\ldots\ldots\ldots \lambda_e &=& 2\,\pi/(d\,B)\,, \end{array}\right\} \quad \textbf{Envelope.} \tag{3.19}$$

Notice that, while the phase moves faster than the speed of "light" (i.e.: 1), the envelope always moves with a speed $-1 < V < 1$, and has width proportional to $\sqrt{1-V^2}$.

Finally, in case you are familiar with the notion of group speed*, notice that (for the linearized Sine-Gordon equation: $\theta_{tt} - \theta_{xx} + \theta = 0$) we have:* (group speed) = 1/(phase speed) *— which is exactly the relationship satisfied by $c_e = V$ and $c_p = 1/V$ for a breather. This is because, for $|x|$ large, the breathers must satisfy the linearized equation. Thus the envelope must move at the group velocity corresponding to the oscillations wave-length.*

Remark 6.3.5 Pseudo-spectral Numerical Method for the Sine-Gordon Equation.
Here we will give a rough idea of a numerical method that can be used to solve the Sine-Gordon equation. This remark will only make sense to you if you have some familiarity with Fourier Series for periodic functions.

The basic idea in spectral methods is that the numerical differentiation of a (smooth) periodic functions can be done much more efficiently (and accurately) on the "Fourier Side" — since there it amounts to term by term multiplication of the n^{th} Fourier coefficient by in. On the other hand,

non-linear operations (such as calculating the square, point by point, of the solution) can be done efficiently on the "Physical Side".

Thus, in a numerical computation using a pseudo-spectral method, all the operations involving taking derivatives are done using the Fourier Side, while all the non-linear operations are done directly on the numerical solution. The back-and-forth calculation of Fourier Series and their inverses is carried by the FFT (Fast Fourier Transform) algorithm — which is a very efficient algorithm for doing Fourier calculations.

Unfortunately, a naive implementation of a spectral scheme to solve the Sine-Gordon equation would require **periodic** *in space, solutions. But we need to be able to solve for solutions that are* **mod-**2π **periodic** *(such as the kinks and anti-kinks), since the solutions to the equation are angles. Thus, we need to get around this problem.*

In a naive implementation of a spectral method, we would write the equation as

$$\left.\begin{array}{rcl} u_t &=& v\,, \\ v_t &=& u_{xx} - sinu\,, \end{array}\right\} \tag{3.20}$$

where $u = \theta$ *and* $v = \theta_t$*. Next we would discretize space using a periodic uniform mesh (with a large enough period), and would evaluate the right hand side using FFT's to calculate derivatives. This would reduce the P.D.E. to some large O.D.E., involving all the values of the solution (and its time derivative) at the nodes in the space grid. This O.D.E. could then be solved using a standard O.D.E. solver — say, ode45 in MatLab.*

In order to use the idea above in a way that allows us to solve the equation with mod-2π *periodicity in space, we need to be able to evaluate the derivative* u_{xx} *in a way that ignores jumps by multiples of* 2π *in* u*.* The following trick works in doing this:

Introduce $\boxed{U = e^{iu}.}$ *Then*

$$u_{xx} = i\,\frac{(U_x)^2 - U\,U_{xx}}{U^2} \tag{3.21}$$

gives a formula for u_{xx} *that ignores* 2π *jumps in* u*.* Warning: *In the actual implementation one must use*

$$u_{xx} = -\mathsf{imag}\left(\frac{(U_x)^2 - U\,U_{xx}}{U^2}\right)$$

to avoid small imaginary parts in the answer (caused by numerical errors).

6.4 Suggested problems.

A list of suggested problems that go along with these notes follow:

1. Check the derivation of the system of equations (2.20).
2. Derive the continuum equation in (2.21).
3. Look at the end of section 6.2, under the title "General Motion: String and Rods". Derive continuum equations describing the motion (in the plane) of a string without constraints.

4. Look at the end of section 6.2, under the title "General Motion: String and Rods". Add bending springs to the model, and derive continuum equations describing the motion (in the plane) of a rod without constraints.

5. Do the check stated in remark 6.3.2.

6. Answer the question in remark 6.3.3.

7. Do the dimensions check stated below equation (3.11).

8. Answer the question in remark 6.3.4.

9. Show that (3.13) is a solution (there is a hint about how to do this a few lines below the equation).

10. Use a computer to plot the solution in (3.13), as a function of z, for a few choices of c.

11. Show that (3.17) is a solution.

12. Use a computer to plot the solution in (3.17), as a function of x, for various times and choices of parameters.

13. Implement a numerical code to calculate interactions of kinks, breathers, etc., using the ideas sketched in remark 6.3.5.