1 Modeling and Inverse Problems

In this chapter we will present a simple application to illustrate the iterative modeling process that was given in Chapter 1. In addition, using this illustrative example the notion of an inverse problem will be discussed. The inverse or parameter estimation problem plays an indispensable role in developing mathematical models for biological and physical systems. As we shall see, because so many different mathematical models are plausible for a given system, model validation is an essential part of the modeling process. Indeed as we formulate the physical or biological problems mathematically, we find that the problem amounts to that of determining one or more unknown parameters in the mathematical model from some (limited) knowledge about the behavior of the system. Problems of this type arise in many important applications including geophysics, ecology, flexible structures, medical imaging and materials testing.
2 Mechanical Vibrations

To begin, we consider a spring of length $l$ attached to a rigid horizontal support (e.g., ceiling) and a small object of mass $m$ hanging from the bottom of the spring (see Figure 1). Now note that if we pull down or push up on the body a distance $\Delta l$, the elastic spring will exert a restoring force to pull the object back up or to push the object down, respectively. If $\Delta l$ is small compared to the spring natural length $l$, then the spring restoring force, denoted by $F_r$, can be described by Hooke’s law (see, e.g., [107, 111]). Mathematically, we write

$$F_r = -k\Delta l,$$

where $k$ is called the spring-constant, which is a measure of the stiffness of the spring. Note that if $\Delta l$ is positive, then the restoring force is negative, whereas if $\Delta l$ is negative, then $F_r$ is positive.

![Figure 1: Spring-mass system (with the mass in equilibrium position).](image)

In modeling the motion of the mass $m$, it will be convenient to describe the position of the mass with respect to its equilibrium position. The equilibrium position of the mass is that point where the mass will hang at rest when no external forces (other than gravity) are being applied. We let $y = 0$ denote this equilibrium point and take the downward direction to be positive. Newton’s Second Law of Motion is fundamental to the description of the position of the mass at time $t$; this states

$$F = ma,$$

where $F$ is the sum of all forces exerted on the mass, $m$ is the body’s mass, and $a$ is the acceleration of the body. Let $y(t)$ denote the position of the mass at time $t$. Using
Newton’s Second Law of Motion, we obtain (see for instance, [104, 105])

\[ m\ddot{y} = -ky. \]  

The differential equation (3) is a second-order, linear differential equation with constant coefficients. Its solution can be readily obtained as (see, e.g., [104, 110])

\[ y(t) = A\cos \omega t + B\sin \omega t, \]  

where \( \omega = \sqrt{k/m} \). The constants of integration \( A \) and \( B \) are determined from the initial conditions, \( y(0) = y_0 \) and \( \dot{y}(0) = v_0 \), and are given by

\[ A = y_0, \]  
\[ B = \frac{v_0}{\omega}. \]

In order to analyze the solution (4), it is convenient to rewrite it as a single cosine function of the form

\[ y(t) = R\cos(\omega t - \phi), \]

where \( R = \sqrt{A^2 + B^2} \) and \( \phi = \tan^{-1}(B/A) \). This solution is depicted in Figure 2. Note that the solution \( y(t) \) lies between \(-R\) and \(+R\), and that the motion of the body is periodic with a period of \( 2\pi/\omega \). This type of motion is called simple harmonic motion, \( \omega = \sqrt{k/m} \) is called the natural frequency of the system, \( R \) is the amplitude of the motion, and \( \phi \) is called the phase angle of the motion.

\[ \begin{array}{c}
\text{Figure 2: Graph of the simple harmonic motion, } y(t) = R\cos(\omega t - \phi). \end{array} \]

In summary, our modeling process as discussed in Chapter 1 begins with the real physical model (a weight (mass) hanging from the bottom of an elastic spring) (step (i)) and
proceeds with force balancing (Newton’s Second Law of Motion) (step (ii)) to derive a mathematical model in terms of a differential equation (step (iii)). We next obtain the analytical solution to our differential equation model (step (v)). However, in comparison to real physical systems of mechanical vibration (step (vi)), the oscillations do not persist over time but eventually die out. This leads us to step (vii) which requires a re-examination in our understanding of the mechanical vibration system. Perhaps, we have over simplified our assumptions. For example, are there other forces (in addition to the spring restoring force) being exerted on the body in Newton’s Second Law of Motion (2)? Specifically, consider a new experiment where we now add to the mass two light “massless” paddles (see Figure 3). As the body moves through the air, there is an apparent resistive force to motion (the paddles are bending in the direction opposite to motion). Furthermore, more bending will occur as the mass moves faster. Simply stated, this force is proportional to the magnitude of the velocity $\dot{y}$ and can be modeled by

$$F_d = -c\dot{y},$$

where $c$ is the viscous damping coefficient. The resistive force $F_d$, which the medium exerts on the body $m$, is also called the damping, or drag force. If we take this new force into consideration, our new mathematical model becomes

$$m\ddot{y} = -ky - c\dot{y}. \quad (9)$$

or

$$m\ddot{y} + c\dot{y} + ky = 0. \quad (10)$$

If we assume that $c^2 - 4km < 0$, every solution of (10) has the form

$$y(t) = e^{-ct/2m}[A \cos \nu t + B \sin \nu t], \quad (11)$$

where $\nu = \sqrt{4km - c^2}/2m$ and $A$ and $B$ are constants to be determined from initial conditions as earlier. If we use similar arguments to those in the undamped case, the damped solution (11) can also be rewritten in the form

$$y(t) = Re^{-ct/2m} \cos(\nu t - \delta). \quad (12)$$

Observe that the solution oscillates between the curves $\pm Re^{-ct/2m}$. That is, the motion of the mass is periodic with decreasing amplitude, as depicted in Figure 4.
Thus, with damping present in the system, the motion of the body always dies out eventually. Engineers usually refer to such systems as spring-mass-dashpot systems. Spring-mass-dashpot systems are ubiquitous in engineering, science, and indeed in nature. For example, they are used as shock absorbers in vehicles to damp out bumps on the road as well as to minimize the recoil effect of a heavy gun barrel. They also find modeling applications in muscle mechanics and molecular level phenomena in materials (e.g., polarization, “electron cloud” models, in response to alternating electric fields).

3 Inverse Problems

Mathematical models as described by equations (3) and (10) above are of the “forward” type; that is, the parameters $m$, $c$, and $k$ are assumed to be known, as well
as the initial conditions. The mathematical model then predicts the resultant model behavior $y(t)$ at any time $t$ from the solution formulas (4) or (11). This is typically the approach taken in sensitivity investigations, which is quite useful, and can provide important features of the model as functions of parameters (see [102, 106, 109, 112] and the references therein). However, in reality, not all parameters are directly measurable (e.g., most springs in mechanical devices come without specification of the spring constant $k$). Instead, we may have sparse and noisy measurements of displacements (using proximity sensors) and/or accelerations (using accelerometers). From this information, we need to find the unknown parameters. Problems of this type are called inverse or parameter estimation problems and are ubiquitous in modeling. Finding the solutions to an inverse problem is, in general, nontrivial because of non-uniqueness difficulties that arise. This undesirable feature is often due to noisy data and insufficient number of observations. For a discussion on the non-uniqueness as well as other issues such as stability in inverse problems we refer the interested reader to [101, 103].

To discuss the inverse problem formulation for the spring-mass-dashpot system, we assume that all three parameters $m$, $c$, and $k$ are unknown and that displacement observations $y^d_i$ at selected temporal points $t_i$ are available. If we have noise free observations (which is never the case in practice), then we only need three well-chosen points $t_i$ to obtain three equations to solve for three unknowns $m$, $c$, and $k$. However, due to noise in the measurements, we usually take $n$ observations. Then, a typical inverse or estimation problem involving (10) is to find $q \in Q_{AD} = \{(m,c,k)|0 < m < M, 0 < c, 0 < k\}$ by minimizing the least squares criterion

$$J(q) = \sum_{i=1}^{n} \left| y_{\text{mod}}(t_i; m, c, k) - y^d_i \right|^2 = \sum_{i=1}^{n} \left| f(t_i; q) - y^d_i \right|^2 \quad (13)$$

Here $y_{\text{mod}}(t_i; m, c, k)$ is the solution to (10) corresponding to $m$, $c$, and $k$. The above procedure leads to a constrained optimization problem. We also remark that such problems also require one to solve for the solution of the differential equation model (10) multiple times.

**MODELING: Formulation of $f(t_i; q)$ = observations of model response**
4 Modeling as an Iterative Process

We give a brief discussion of certain philosophical notions that are important in the modeling of physical and biological systems. Modeling in our view is simply a means for providing a conceptual framework in which real systems may be investigated. The modeling process itself is (or should be) most often an iterative process: one can distinguish in it a number of rather separate steps that usually must be repeated. This iterative modeling process is schematically depicted in Figure 5. One begins with the real system under investigation and pursues the following sequence of steps:

(i) empirical observations, experiments, and data collection;

(ii) formalization of properties, relationships and mechanisms that result in a biological or physical model (e.g., stoichiometric relations detailing pathways, mechanisms, biochemical reactions, etc., in a metabolic pathway model; stress-strain, pressure-force relationships in mechanics and fluids);

(iii) abstraction or mathematization resulting in a mathematical model (e.g., algebraic and/or differential equations with constraints and initial and/or boundary conditions);

(iv) formalization of uncertainty/variability in model and data resulting in a statistical model (this usually involves basic assumptions about errors in modeling, observation process/ measurement, etc.);

(v) model analysis that can consist of simulation studies, analytical and qualitative analysis including
stability analysis, and use of mathematical techniques such as perturbation studies, parameter estimation (inverse problems) data fitting, statistical analysis;

(vi) interpretation and comparison (with the real system) of the conclusions, predictions and conjectures obtained from step (v);

(vii) changes in “understanding” of mechanisms, pathways, etc., in the real system.

As one completes step (vii), one is led naturally to reformulate the physical or biological model by returning to either step (i) (if new experiments are indicated) or step (ii). In either case one then proceeds through the steps again, seeking to improve the findings of the previous transit through the sequence.

Steps (i), (ii), (iii), (iv) belong to what one might term the formulation stage of the modeling process, while step (v) is the solution stage of the modeling process, and steps (vi) and (vii) constitute the interpretation stage. In practice, however, it is often (unfortunately) the case that investigators do not make a clear distinction in the steps outlined here. This can lead to confusion and, in some cases, incorrect conclusions and gross misunderstanding of the real system.

Let us turn next to the reasons frequently given for modeling.

Perhaps the one most often offered is simplification: the use of models makes possible the investigation of very complex systems in a systematic manner. A second rationale is ease in manipulation: investigations involving separation of subunits and hypothesis testing may often be facilitated through use of simulations in place of experimentation. The suggestive features in modeling can also help in formulation of hypotheses and in the design of critical experiments. The modeling process also requires preciseness in investigation in that one must move from a general, verbal explanation of phenomena to a specific, quantitative one.

But a rationale perhaps more fundamental than any of these is that modeling leads to an organization of inquiry in that it tends to polarize one’s thinking and aid in posing basic questions concerning what one does and does not know for certain about the real system. Whatever the reasons that have been advanced to justify modeling attempts, it is sufficient perhaps to note that the primary goal must be enlightenment,
The Iterative Modeling Process

(i) Empirical Observations (experiments and data collection)
(ii) Formalization of properties, relationships and mechanisms which result in a biological or physical model
(iii) Abstraction or Mathematization resulting in a mathematical model
(iv) Formalization of Uncertainty/Variability in model and data resulting in a statistical model
(v) Model Analysis
(vi) Interpretation and Comparison (with the real system)
(vii) Changes in understanding of mechanisms, etc., in the real system.

Formation Stage: (i),(ii),(iii),(iv) Solution Stage: (v) Interpretation Stage: (vi), (vii)

Figure 5: Schematic diagram of the iterative modeling process.

that is, to gain a better understanding of the real system, and the success or lack thereof of any modeling attempt must be appraised with this in mind.

One must recognize the various levels or multi-scale aspects of modeling in any attempt to compare or assess the validity of several models for a phenomenon. For example, consider the phenomena involved in the transmission of a nerve impulse along an axon: this process is likely to be described by the mathematician or biophysicist in terms of partial differential equations, wave phenomena, or transmission line analogies, whereas a neurophysiologist might speak in terms of local circuit analogies and changes in conductances. The cell physiologist might describe the phenomena in the context of transport properties of membranes and ion flow, while the molecular biochemist could insist that the real story lay in the theory of molecular binding.

A second example involves the physical motion (vibration) of a structure such as a plate or beam. Again the mathematician might describe this in terms of a partial differential
equation whereas the mechanical engineer might use a modal analysis (in terms of natural frequencies of oscillation) based on internal stress-strain relationships. In each of the examples cited above, the different modeling approaches move to an increasingly more micro level. Each approach involves an attempt to explain a phenomenon that is not understood at one level by description at a more micro level (in general) where understanding is more complete. This attempt to explain “unknowns” in terms of more basic “knowns” is clearly the foundation of most modeling investigations. Indeed, in addition to noting that nerve impulse phenomena are described in terms of membrane conductances, permeabilities, ion flow, etc., one might observe that blood circulation is studied in the context of elementary hydrostatics and fluid dynamics while metabolic processes are usually investigated via use of the language of elementary chemical kinetics and thermodynamics.

The choice of the level (micro vs. macro) at which one models depends very much upon the training and background of the investigator. Furthermore, the perception of whether a model is a “good” one or not is also greatly influenced by this factor, and it is therefore not surprising that all of the approaches to the nerve impulse phenomena mentioned above (or indeed those for modeling any physical or biological phenomena) can be subjected to valid criticisms in any attempt to evaluate them.

Before discussing the criteria one might use in evaluating modeling investigations, let us list some of the common difficulties and limitations often encountered in the modeling of systems:

(a) **Availability and accuracy of data;**

(b) **Analysis of the mathematical model;**

(c) **Use of local representations that are invalid for the overall system;**

(d) **Obsession with the solution stage;**

(e) **Assumption that the “model” is the real system;**

(f) **Communication in interdisciplinary efforts.**

The first item in this list requires no further comment; the second includes both theoretical and computational difficulties in the mathematical treatment of a given set of equations. Although formidable obstacles can still arise, this is a much less critical problem today in modeling than it was, say, in the physical sciences in Newton’s time. This is due in large part to great strides that have been made in the last several decades with
the advance of modern computing facilities and the concomitant development of rather sophisticated numerical procedures. We remark that (c) is especially prevalent in certain physiological modeling, where systems are not easily manipulated experimentally. In vitro data and parameter values (determined via experimentation in non physiological ranges) are often used to model, predict and draw conclusions about in vivo situations. While (d) is likely to be a problem for investigators with a mathematical or physics background (in their enthusiasm for finding solutions of their model equations and various generalizations, they tend to forget or ignore the fact that the model is only an approximation and that certain aspects of the physical or biological model on which it is based are very poorly understood), item (e) can be a problem for both mathematical and physical and/or biological scientists. Even physicists and biologists sometimes have a penchant for disbelieving data that contradicts model simulations and predictions. It can be very tempting to throw out “faulty” data rather than reformulate the basic model. Finally, because most serious physical and biological modeling projects involve an interdisciplinary effort, there is always the possibility of serious lack of communication and cooperation due to differences in vocabulary, goals, and attitudes. Often mathematicians are only looking for a “problem” to which their already highly developed theories and techniques apply; i.e., they are in possession of a “solution” and in search of the “problem” they have solved! On the other hand, physicists and biologists can be too impatient with the mathematicians’ desire to hypothesize rather implausible mechanisms and relationships (which can sometimes lead to exciting new perspectives about a phenomenon!)

Finally, we turn to the question of how one appraises a specific modeling attempt. There are a number of criteria that one might use. Among those proposed by various authors are the suggestions that a good model should: fit data accurately; be theoretically consistent with the real system; have parameters with physical meaning that can be measured independently of each other; prove useful in prediction; not so much explain or predict, but organize and economize thinking; pose new empirical questions and help answer them through the iterative process; help us understand the phenomena it represents and think comfortably about them; and point to inadequacies in some way of available data. It is clear, though, that for a modeling investigation to be deemed a success, it must have enhanced our overall knowledge and understanding of the phenomena in question. As one of our students (having been attacked by other students for some rather unorthodox and, at the time, unsupported hypothesis about mechanisms) noted in defending his efforts, “We learn little indeed if the models we build never stretch our understanding, but only tell us what we already feel is safely known.” We remind the reader of the often quoted truth “all models are incorrect, but some are more useful than others”.

In concluding our philosophical remarks, we remark that one can distinguish between at least two basic types of scientific models: descriptive and conceptual models. Descriptive models, those designed to explain observed phenomena, will be the focus of our attention here. Conceptual models, models constructed to elucidate delicate and difficult points
in some scientific theory, are often used to help resolve apparent paradoxes involving two
descriptive models. Conceptual models do not appear widely in the biological literature
since in many cases basic descriptive models are still under development.
5 MODELS USUALLY BASED ON CONSERVATION LAWS

- Force and Momentum Balance (mechanics, physical models)
- Mass Balance (biological, chemical)
- Energy Balance (thermal)
6 Mass Balance and Mass Transport

7 Introduction

Mass transfer is important in many areas of science and engineering. Many familiar phenomena involve mass transfer:

- The spreading of odorous gas in a room.
- Liquid in an open pail of water evaporating into surrounding air.
- A piece of sugar added to a cup of coffee eventually dissolving by itself into the surrounding solution.
- Transport of chemical substances into the red blood cells.
- Transport of O$_2$ throughout the human body - systemic and cellular.

The most elementary approach to mass transport is compartmental analysis. Compartmental modeling has been and is being used widely in many branches of biology, biomedicine, and in pharmacokinetics as well as in physical modeling. Indeed, one can find examples of compartmental modeling in almost any publication of the major journals in physiology and pharmacology. In addition, there are several books that cover both the theory and applications of compartmental modeling, e.g., [3, 4], while several books have chapters giving introductions to compartmental analysis as well as its applications (see for instance, [1, 5, 6, 7, 8]).

8 Compartmental Concepts

A compartment is an abstraction used often in biological (and other scientific) models. It may of course be a physical entity, a distinct space having discernible boundaries across which material (energy) moves at a measurable rate (and for
which, as a rule, an “inside” and “outside” are readily distinguishable). More generally, we might take as a compartment any anatomical, physiological, chemical, or physical subdivision of a system throughout which the behavior (e.g., concentration) of a given substance is uniform. It can also be useful to compartmentalize in terms of different types of molecules or chemical forms (e.g., hemoglobin, red blood cells, blood plasma). We might then make a formal definition of a compartment as follows: if a substance $S$ is present in a system in several distinguishable forms or locations and if $S$ passes from one form or location to another form or location at a measurable rate, then each form or location constitutes a separate compartment for $S$.

The compartment concept represents a system as a set of interconnecting components or subsystems. We further remark that the compartments (subsystems) do not always correspond to physically identifiable components. A couple of very simple examples serve to illustrate this concept. In studying certain diseases, it is convenient to regard each stage of the disease as a compartment and to construct a mathematical model based on the transfer between them. Another common example of tracer studies involves red blood cells suspended in an isotonic (uniform tension or osmotic pressure) fluid. In this case one might be interested in the concentration of radioactive potassium ions in the million of separate physical compartments. However, for modeling uptake phenomena, it is most likely that one would consider the collection of red blood cells as a whole and formulate a two-compartment model consisting of a fluid compartment and a red blood cell “compartment”.

These examples illustrate the fact that it is the behavior of a substance $S$ in a system which determines the compartmentalization of the system and not necessarily the physical situation itself. Differences in how investigators perceive this “behavior” often lead to the dramatically different compartmentalizations of a given system found in the literature. For an example, one might be surprised at the wide range of models used to describe the glucose homeostatic system in mammals.

In the modeling of mass transport between compartments, several assumptions are commonly made. Among these are:

(i) constant-volume compartments,

(ii) well-mixed compartments, and

(iii) for systems in which transport is across a mem-
brane, constancy of the transport coefficient $K$ (discussed further below) in time.

Whether any or all of these assumptions can be justified depends very much on the nature of the phenomena and systems being modeled. While a decision to posit (i) is usually rather straightforward, support of (ii) is often more difficult. There are a number of major contributors to rapid distribution within a compartment, including

(a) stirring or mixing by currents within the body of the solution,

(b) transportation (convection) by a flowing stream, and

(c) diffusion (thermal motion of solute molecules).

Contributions to well-mixing by (a) and (b) can be valid even when the distances (compartment size) are substantial, while (c) is usually a valid component only in the case of small-volume compartments.

The convenience of this type of decomposition (compartmentalization) is that it leads directly to a set of equations based on simple balance relations. This can be stated simply as:

$$\text{change in compartment } j = (\text{sum of all transfers into compartment } j) - (\text{sum of all transfers out of compartment } j) + (\text{creation within compartment } j) - (\text{destruction within compartment } j).$$

9 Compartment Modeling

To illustrate the ideas behind the concept of a compartment and how it is used, we discuss the simplest example of a two compartment model. Consider two chambers separated by a membrane with solute $S$ and water in each chamber (see Figure 6). Assume that each chamber is well-mixed (or well-stirred); that is, when the solute $S$ is added to the water it is instantly distributed throughout the chamber. This process is slower for liquid than gas and is slowest for solid (it can be achieved by mixing or by convection by a stream.)
Naturally, one is faced with the following questions. What are the compartments? How many? The answers depend very much on how the solute behaves in the system.

- **If the membrane is highly permeable** (full of holes), one compartment is adequate to describe the concentration of solute. In this case, equilibration is essentially instantaneous.

- **If the membrane is impermeable** (no transport across membrane occurs), only one compartment (the one to which solute is added) is needed to model the solute concentration.

- **If the membrane is permeable**, two compartments are needed and transport of solute between the compartments must be modeled.

In addition, in the modeling of mass transport between compartments separated by a membrane, the above assumptions (i)-(iii) are usually made. The basic parameter involved in membrane separated compartmental exchange is called the *transport coefficient*. It is usually denoted by $K$ and is proportional to physical properties of the
compartments. In particular,
\[ K \propto \frac{A}{\delta} \]
\[ = c \frac{A}{\delta}, \]
where \( c \), the proportional constant, is called the membrane permeability coefficient with units of \( \frac{m^2}{sec} \), \( A \) is the cross-sectional area (in units \( m^2 \)) and \( \delta \) is the thickness of the membrane (in \( m \)). This implies that \( K \) has units of \( \frac{m^2}{sec} \), which is the rate at which a substance (volume) is transported across the membrane. This is also sometimes called the \textit{volumetric rate}.

Recall the definition of mass density given by
\[ \rho = \text{mass density (or mass concentration)} = \frac{\text{mass of solute}}{\text{volume of solution}} \text{ in } \frac{kg}{m^3}. \]

By letting \( V \) denote the volume of the compartment into or out of which we are modeling solute transport, we can now write
\[
\left\{ \text{rate of change of mass in compartment} \right\} = \left\{ \text{volumetric rate} \right\} \times \{\text{mass density}\}.
\]

Then, we have
\[ \frac{d}{dt}m = K \rho. \]

We next consider using these concepts in a two compartment model such as depicted in Figure 6. In this formulation we assume:

- A two-compartment system labeled 1 and 2 with constant volumes \( V_1 \) and \( V_2 \).
- A solute is present and is transported between compartments across the membrane with transport coefficients \( K_{1,2} \) (from 1 to 2) and \( K_{2,1} \) (from 2 to 1), which for the moment are not assumed to be equal. The masses of solute in compartments 1 and 2 are denoted by \( m_1 \) and \( m_2 \) respectively.

Simple mass balance considerations in compartment 1 lead to the following differential equation:
\[
\frac{d}{dt}m_1 = (\text{rate of transfer into 1}) - (\text{rate of transfer out of 1})
\]
\[ = K_{2,1} \rho_2 - K_{1,2} \rho_1. \]
Similarly, for compartment 2 we obtain
\[
\frac{d}{dt}m_2 = K_{1,2}\rho_1 - K_{2,1}\rho_2.
\]
We may rewrite this in terms of concentrations (or densities) by using \( \rho_i = \frac{m_i}{V_i} \) to obtain
\[
\frac{d}{dt}\rho_1 = \frac{1}{V_1}[K_{2,1}\rho_2 - K_{1,2}\rho_1]
\]
\[
\frac{d}{dt}\rho_2 = \frac{1}{V_2}[K_{1,2}\rho_1 - K_{2,1}\rho_2].
\]
From the above calculations, we observe the following important consequences:

- If we assume \( K_{1,2} = K_{2,1} = K \), then
  \[
  \frac{dm_1}{dt} = K(\rho_2 - \rho_1)
  \]
  \[
  \frac{dm_2}{dt} = K(\rho_1 - \rho_2)
  \]
  \[
  = -\frac{dm_1}{dt}
  \]
  by laws of mass conservation. If \( \rho_2 > \rho_1 \), then \( \frac{dm_1}{dt} > 0 \) (that is, the mass of solute in chamber 1 increases due to movement of solute from chamber 2 (high concentration) to chamber 1 (low concentration)).

This type of mass transport is called *passive transport* or *molecular (membrane) diffusion*. (It is very similar to the manner in which heat is transported in a rod – one observer holds one end of a rod and when the other end is heated, the part that is held will become hotter even though it is not in direct contact with the heat source; thus, heat is said to be transported (conducted) from high concentration or temperature to low concentration or temperature.)

- In terms of mass *concentrations* (or mass *densities*) we have
  \[
  \frac{d\rho_1}{dt} = \frac{K}{V_1}(\rho_2 - \rho_1)
  \]
  \[
  \frac{d\rho_2}{dt} = \frac{K}{V_2}(\rho_1 - \rho_2).
  \]
  Note that \( \rho_1 \neq -\rho_2 \) unless \( V_1 = V_2 \). That is, in general, we do not have concentration balance.
It is important to note: We have mass conservation and not concentration (or density) conservation.

There are several advantages as well as disadvantages that arise when using simple compartmental models.

Advantages:

It is relatively straightforward to write down the mass balance equation (input – output relation). In addition, the resulting model is a set of ordinary differential equations (often rather easy to solve analytically or numerically).

Disadvantages:

The solution is assumed to be well-mixed. To see the inherent limitations, we can consider, for example, dropping a blue liquid dye into a bucket of water. The dye will diffuse slowly to other parts of the water. That is, the concentration of the dye is different in different parts of the bucket. Often, one can satisfy the well-mixed assumption by considering very small volumes or by stirring the compartment. However this is not always reasonable. For example, the transport of drug in the liver will have different concentrations through different parts of the liver. The concentration of a drug injected into the systemic blood may have different concentrations in different parts of the blood circulation system.

10 General Mass Transport Equations

Recall from compartment analysis, we have

\[ \frac{dm_1}{dt} \propto (\rho_2 - \rho_1), \]

that is, the rate of change of mass is proportional to concentration difference. This type of transport process is known as molecular diffusion. To illustrate this concept, consider the movement of individual molecules, say A and B, in a fluid as depicted in Figure 7. Suppose that there are more A molecules near region (1) than near region (2) and since molecules move randomly in both directions, more A molecules will move from (1) to (2) than from (2) to (1). The net transport of A is from a high concentration region to a low concentration region; this is molecular diffusion.

We further remark that:

- As molecules move they change directions by bouncing off other molecules after collisions. Since they travel in a random path, molecular diffusion is also called a random walk process.
To increase the rate of mixing of a substance in solution, the liquid can be mechanically agitated by a device and **convective mass transfer** will occur (due to movement of the bulk liquid).

Let us now consider a mixture of several species (labeled with index $i$) in a moving fluid through a pipe as depicted in Figure 8.

We will formulate a mass balance relationship for species $i$ on a volume element of thickness $\Delta x$ as shown. The general mass balance on species $i$ is

$$
\begin{pmatrix}
\text{rate of accumulation of mass } i \\
\text{in volume element}
\end{pmatrix}
= 
\begin{pmatrix}
\text{rate of mass } i \\
\text{entering face } x
\end{pmatrix}
- 
\begin{pmatrix}
\text{rate of mass } i \\
\text{leaving face } x + \Delta x
\end{pmatrix}
\pm 
\begin{pmatrix}
\text{rate of generation (or consumption) of mass } i \\
\text{(by metabolism or chemical reaction)}
\end{pmatrix}
$$

To write down the rate of mass entering and leaving, we need to discuss flux laws for mass transport. (Mass flux is defined as the mass that passes through a unit cross sectional area per unit time.) We do this first in the case in which the carrier fluid itself is stationary, that is, the fluid bulk velocity $v$ is zero.
10.1 Mass Flux Law in a Stationary (Non-Moving) Fluid

Since we are dealing, in general, with multiple species, the “concentrations” of the various species may be expressed in numerous ways. We begin by defining mass density (or mass concentration) at a point \( p = (x, y, z) \) by

\[
\rho(t, x, y, z) = \frac{dm}{dV} = \lim_{\Delta V \to 0} \frac{1}{\Delta V} \int_{\Delta V} m(t, \tilde{x}, \tilde{y}, \tilde{z}) \, dV,
\]

where \( \Delta V \) is a small element of volume containing the point \( p \) with \( m(t, \tilde{x}, \tilde{y}, \tilde{z}) \) being the mass of the particle located at \( (\tilde{x}, \tilde{y}, \tilde{z}) \in \Delta V \). We make the following assumptions for our derivation.

(i) In the small volume element \( \Delta V = \Delta x A \) (see Figure 9), we have well mixing so that \( \rho \) is constant in \( \Delta V \).

(ii) Species are uniform in \( y \) and \( z \) directions (that is, \( \rho = \rho(t, x) \)).

Figure 9: Incremental volume element.
In a diffusing mixture involving multiple species, the various chemical species may be moving at different velocities. Let \( v_i \) denote the velocity of species \( i \) with respect to a stationary coordinate system. Then we may define the local “mass average velocity” by

\[
\bar{v} = \frac{\sum_{i=1}^{n} \rho_i v_i}{\sum_{i=1}^{n} \rho_i}.
\]

In some cases one is interested in the velocities of a given species \( i \) relative to \( \bar{v} \) (or perhaps some other velocity) rather than relative to the stationary coordinate system \( (v_i) \). This leads to the definition of the “relative diffusion velocities” \( v_{ir} \) given by

\[
v_{ir} = v_i - \bar{v} = \text{diffusion velocity of } i \text{ relative to } \bar{v}.
\]

We may use the mass balance for species \( i \) in the element of volume \( \Delta V = A \Delta x \) with cross sectional \( A \) (which may depend on \( t \) and/or \( x \)). If we assume no creation or destruction of mass for the present, we may define \( \dot{q}_i \) by

\[
\dot{q}_i = \text{rate of mass transport of species } i \text{ (with mass concentration } \rho_i).\]

We use the compartmental analysis techniques, treating the element of volume \( \Delta V \) as a “thin membrane” between the immediate “compartments” where the concentrations are \( \rho(x) \) and \( \rho(x + \Delta x) \), respectively. We find that \( \dot{q}_i \) is proportional to \( A \frac{\Delta \rho}{\Delta x} \). Then we may write

\[
\dot{q}_i = AD_i \frac{\rho_i(x) - \rho_i(x + \Delta x)}{\Delta x},
\]

where the constant of proportionality is given by \( D_i \) and is called the mass diffusivity constant (in units \( \frac{m^2}{sec} \)).

Note that here we have assumed that \( A \) is approximately constant for the small volume. The above expression is the rate of mass transport in the incremental volume element. To find the rate of mass transport at an arbitrary point \( x \), we let \( \Delta x \to 0 \) to obtain

\[
\dot{q}_i \to -AD_i \frac{\partial \rho_i}{\partial x}.
\]

Recall mass flux for species \( i \) is \( j_i = \frac{\text{rate of mass transport}}{\text{cross sectional area}} \). Hence, we have

\[
j_i = -D_i \frac{\partial \rho_i}{\partial x} \text{ with units } \frac{kg}{m^2 \cdot sec}.
\] (14)

The following remarks are in order:
1. This is known as **Fick’s first law of diffusion** [2], which says that mass flux is proportional to the mass concentration gradient; in general, temperature, pressure gradients, and external forces also affect the flux, but their effects are usually minor and are ignored, or else treated through dependence of the diffusion coefficient $D_i$ on them.

2. We will later see that Fickian diffusion is very similar to Fourier’s law of heat conduction and Newton’s law of momentum (in one-dimensional problems).

3. The negative sign in (14) agrees with the observation that mass flows from high to low mass concentration. If we have $\rho_i(x) < \rho_i(x + \Delta x)$, we find that

$$\frac{\partial \rho_i}{\partial x} > 0$$

and hence net flow is in the opposite direction from the positive x-direction.

4. In three-dimensional problems, these concepts all readily generalize, and for mass density $\rho_i(t, x, y, z)$, we find that the mass flux is given by

$$\vec{j}_i = -D_i \nabla \rho_i.$$  

5. The mass flux with respect to the stationary coordinates is given by

$$j_i = \rho_i v_i,$$

and the mass flux with respect to the relative diffusion velocity $j_{ir}$ is given by

$$j_{ir} = \rho_i v_{ir}.$$

### 10.2 Mass Flux in a Moving Fluid

We assume that the bulk velocity is denoted by $v$, so the total velocity of species relative to the fixed coordinate system is $v_i = v_{i,\text{diff}} + v$, and hence the total flux of species $i$ relative to a fixed point in the stationary coordinate system is $j_{i,\text{tot}} = \rho_i v_i = \rho_i v_{i,\text{diff}} + \rho_i v = j_{i,\text{diff}} + j_{i,\text{bulk}}$, where we recall the diffusive flux was given by

$$j_{i,\text{diff}} = -D_i \frac{\partial \rho_i}{\partial x}.$$  

Hence, $j_{i,\text{tot}} = -D_i \frac{\partial \rho_i}{\partial x} + \rho_i v$.

Now write the mass balance on a small element:

$$\frac{\partial}{\partial t} [\rho_i \Delta x A(t, x)] = j_{i,\text{tot}} A|_x - j_{i,\text{tot}} A|_{x+\Delta x} + r_i \Delta x A,$$  

24
where \( r_i \) is rate of production (destruction) of species \( i \) per unit volume. Dividing by \( \Delta x \) and taking the limit as \( \Delta x \to 0 \), we obtain

\[
\frac{\partial}{\partial t}(\rho_i A) = -\frac{\partial}{\partial x}(j_i^{\text{tot}} A) + r_i A \\
= -\frac{\partial}{\partial x}(-AD_i \frac{\partial \rho_i}{\partial x} + A \rho_i v) + r_i A
\]

or

\[
\frac{\partial}{\partial t}(\rho_i A) + \frac{\partial}{\partial x}(\rho_i v A) = \frac{\partial}{\partial x}(AD_i \frac{\partial \rho_i}{\partial x}) + r_i A.
\]

If \( A \) is constant, we obtain the usual mass transport equation

\[
\frac{\partial}{\partial t}(\rho_i) + \frac{\partial}{\partial x}((\rho_i v)) = \frac{\partial}{\partial x}(D_i \frac{\partial \rho_i}{\partial x}) + r_i,
\]

where the second term is identified with the convective or advective transport, and the third term is diffusive transport.

The above derivation can be generalized to the multiple species case to obtain

\[
\frac{\partial}{\partial t}(\sum \rho_i A) = -\frac{\partial}{\partial x}(\sum j_i^{\text{tot}} A) + \sum r_i A,
\]

and since \( \sum r_i = 0 \) (total conservation of mass), \( \sum j_i^{\text{tot}} = \sum \rho_i v_i^{\text{tot}} = \rho v \) (note that the bulk velocity thus agrees with the local mass velocity \( v = \frac{\sum \rho_i v_i}{\sum \rho_i} \) and \( \sum \rho_i = \rho \), we have

\[
\frac{\partial}{\partial t}(\rho) + \frac{\partial}{\partial x}(\rho v) = 0.
\]

This is the well known equation of continuity.

All of the above generalizes to the three-dimensional problem. In particular, the three-dimensional mass transport equation or Diffusion-Convection-Reaction (DCR) equation has the form:

\[
\frac{\partial}{\partial t}(\rho_i) + \nabla \cdot (\rho_i v) = \nabla \cdot (D_i \nabla \rho_i) + r_i.
\]

The equation of continuity in three-dimensions is given by:

\[
\frac{\partial}{\partial t}(\rho) + \nabla \cdot (\rho \vec{v}) = 0. \tag{15}
\]
If \( \rho \) is constant, we obtain \( \nabla \cdot \vec{v} = 0 \). This is known as *incompressibility* of the fluid in which the solute is contained.

**Special cases:**

1. When the bulk velocity, \( \vec{v} \), and the reaction rate, \( r_i \), are both zero, we obtain
   \[
   \frac{\partial}{\partial t} (\rho_i) = \nabla \cdot (D_i \nabla \rho_i),
   \]
   which is called **Fick’s second law of diffusion** or simply the *diffusion equation*.

2. In the case the bulk velocity, \( \vec{v} \), is zero, we have
   \[
   \frac{\partial}{\partial t} (\rho_i) = \nabla \cdot (D_i \nabla \rho_i) + r_i,
   \]
   which is known as the **reaction-diffusion equation**.

3. When the diffusion, \( D_i \), is zero, we have
   \[
   \frac{\partial}{\partial t} (\rho_i) + \nabla \cdot (\rho_i \vec{v}) = r_i.
   \]
   This is known as the **plug-flow**, **ideal tubular**, or **unmixed flow model**. Here the flow of the fluid is orderly with no element of fluid mixing or overtaking (see Figure 10). A necessary and sufficient condition for plug flow is that the residence time is the same for each species.
Exercise: Transport Equations

In the literature one also often finds mass transport in terms of molar concentration $c_i$ and mass fraction $\omega_i$. This exercise will provide experience in deriving mass transport equations in terms of these variables.

(i) Define the molar concentration $c_i$ of species $i$ by $c_i = \rho_i/M_i$, where $\rho_i$ is mass concentration (in units kg/m$^3$) of species $i$ and $M_i$ is the molecular weight (in units kg/moles) so that $c_i$ has units moles of $i$/m$^3$. Define the mass fraction $\omega_i = \rho_i/\rho$, where $\rho$ is the total mass density $\rho = \sum \rho_i$.

Use compartmental analysis to argue that the rate of mass transport at a point $x$ is given by

$$\dot{q}_i = -A\rho D_i \frac{\partial \omega_i}{\partial x}$$

and the mass flux of species $i$ is given by

$$j_i = \rho D_i \frac{\partial \omega_i}{\partial x}.$$

Explain when this is equivalent to Fick’s first law of diffusion.

(ii) Now use this and mass balance principles to derive the general mass transport equations with diffusive and convective terms in terms of the variable $\omega_i$ (as opposed to in terms of $\rho_i$ as done earlier in this chapter).
11 Heat Conduction

12 Motivating Problems

12.1 Radio-Frequency Bonding of Adhesives

Radio-frequency (RF) curing of adhesives is a commercially important process which is used in a number of applications. These include the fixation of prosthetic joints in some fields of medicine, the acceleration of adhesive setting in the woodworking industry, and the bonding of parts in the automotive industry. More specifically, in the automobile industry, the use of non-metallic automotive exterior body panels has grown significantly over the last decade. The most common of these materials is sheet molding compound (SMC), a glass reinforced polyester which provides corrosion resistance, weight reduction, and complex shape molding capability. These parts are typically molded in two layers and adhesively bonded in sandwich fashion around their perimeters to form rigid structures.

The adhesive is commonly applied in a viscous liquid or paste form. Radio frequency, or dielectric, heating is often used to accelerate the cure rate of the adhesive. In this application, the SMC/adhesive/SMC joint is placed between two electrodes (Figure 11). These electrodes then make contact with the joint, compressing it to the desired adhesive bonding thickness. A high voltage electric field, oscillating at approximately 30 MHz, then passes through the joint for a predetermined period of time at preset power levels, exciting polar or ionic species in the adhesive materials and generating heat. In comparison to common adhesives, the SMC is dielectrically relatively inactive. Significant heat can be generated within the adhesive, however, causing it to rapidly undergo a phase transition from liquid to solid (curing), and effectively bonding the two substrates to each other. This process, which can be closely simulated on a laboratory scale using a smaller version of the RF bonding equipment described above, provides us with a physically interesting problem. We must deal with thermally dependent nonlinearities arising from the radio-frequency field itself (i.e., temperature dependent input terms as well as conductivities), and complex internal phase transitions which are parametrized by the degree of cure. This process thus provides us with a problem that is mathematically very interesting. It combines serious modeling issues, mathematical analysis, and computational methodology, while providing a foundation for necessary parameter estimation problems and nonlinear control methodology development.

This industrial problem was a joint collaborative effort between scientists at Lord Corporation (Cary, North Carolina) and faculty and graduate students at North Carolina State University. The goal is to model the radio-frequency curing of epoxy adhesives in bonding of composites. For a detailed development of the mathematical model for the heat transfer through the joint we refer the reader to reference [91]. The model is a version of the “heat equation” of Fourier fame plus terms that take into account the
internal exothermic reaction (which is part of the curing process) as well as the heat generated by the conversion of electrical energy to molecular vibrational energy.

12.2 Thermal Testing of Structures

Recently, associated with the use of fiber reinforced composite materials as well as with more traditional composite metal alloys for aerospace structures, there is growing interest in the detection and characterization of structural flaws (e.g., cracks, delamination, and corrosion) that may not be detectable by visual inspection. An evaluation procedure for such damage detection is of paramount importance in the context of aging aircraft (both civilian and military). One recent effort has focused on nondestructive evaluation (NDE) methods based on the measurement of thermal diffusivity in composite materials (see, e.g., [96]). The idea of this approach is embodied in Figure 12.

In [92] the search for structural flaws in materials is formulated as an inverse problem for a heat diffusion system. From a physical point of view, the system state is the temperature distribution as a function of time and space, the boundary input represents the thermal source (for example, by a laser beam) and the output corresponds to the observation of the temperature distribution at the surface of the material (for example, by an infrared imager), see Figure 12 and [96] for more details. The problem is then of identifying, from input and output data, the geometrical structure of the boundary (i.e., the corroded surface). The mathematical model, which relates front surface temperature (the output data) and back surface “geometry”, is described by the heat equation with appropriate initial and boundary conditions (see [92] for a detailed description).
13 Mathematical Modeling of Heat Transfer

13.1 Introduction

In addition to the two examples discussed in §12, the transfer of energy in the form of heat occurs in numerous industrial production problems including those in the chemical industry, the paper industry, and numerous other production processes. For examples, heat transfer occurs in the drying of lumber, chilling of food and biological materials, combustion problems (burning of fuel), and evaporation processes. In general, heat transfer is energy in transit due to temperature differences and hence “energy balance” is the underlying conservation principle. This transit of energy can occur through conduction, convection, and/or radiation.

- **Conduction.** Conduction generally refers to heat transfer related to molecular activity and may be correctly viewed as the transfer of energy from the more energetic to the less energetic particles of a substance or material due to direct interaction between the particles. This type of transfer is present to some extent in all solids, gases, or liquids in which a temperature gradient exists. It is associated with an empirically based rate formulation known as Fourier’s law to be discussed below. The conduction mode of heat transfer can be related to the random motion of molecules in a gas or substance undergoing no bulk motion or macroscopic movement and is therefore termed diffusion of energy or heat diffusion.

- **Convection.** Heat transfer can also occur in a gas or fluid undergoing bulk or macroscopic motion. The molecules, or aggregates of molecules, move collectively and in the presence of temperature differences, give rise to energy transfer. The molecules retain, of course, their random motion and thus the energy transferred is a superposition of energy transfer due to random motion of particles as well as due...
to bulk motion of the fluid. The cumulative transport is usually called \textit{convection} while the transfer due to bulk motion alone is called \textit{advection}, although this clear distinction is not always made. In modeling, a distinction can be made between forced convective heat transfer, where a fluid is forced to flow past a solid surface by a pump, for example, and natural or free convection which arises most often when a gas or fluid passes over a surface when the two are at different temperatures causing a circulation due to a density difference resulting from the temperature differences in the fluid. For either case, the associated empirical rate “law” is called \textit{Newton’s law of cooling}.

- \textit{Radiation}. Thermal \textit{radiation} refers to energy emitted by matter at a finite positive temperature. This is usually attributed to changes in electron configurations in atoms and molecules that result in the emission of energy via electromagnetic waves or photons and may occur in solids, fluids or gases. The most important example of radiation is the transport of heat to the earth from the sun. The associated quantitative rate “law” is given by the \textit{Stefan-Boltzmann law}.

### 13.2 Fourier’s Law of Heat Conduction

For general molecular transport, all three main types of rate transfer processes - momentum transfer, heat transfer, and mass transfer - are characterized by the same general type of equation. This basic equation is given as follows:

\[
\text{rate of a transfer process} = \frac{\text{driving force}}{\text{resistance}}. \tag{16}
\]

Equation (16) simply states that in order to transfer a property (for example, heat) a driving force needs to overcome a resistance. The transfer of heat by conduction also follows this basic principle and is known as Fourier’s law of heat conduction in fluids or solids. It is written mathematically as

\[
\dot{q} = -kA \frac{\partial u}{\partial x}, \tag{17}
\]

where $\dot{q}$ is the rate of heat transfer and is given in units of power, i.e., watts (W), where $1\text{W} = 1\text{J/sec} = 0.23885$ calories/sec, $A$ is the cross-sectional area normal to the direction of heat flow in m$^2$, $k$ is the \textit{thermal conductivity} in W/m°C, $u$ is the temperature in °K (or °C), and $x$ is the distance in m. This “law” is based on phenomenological or empirical observations (such as a constitutive assumption or “law” in particle mechanics, i.e., Newton’s second “law” of motion, $F = ma$) as opposed to being based on first principles. It satisfies our intuition that the rate of heat transfer across a surface should be proportional to the surface area $A$ and the temperature difference (i.e., the limit of $(u(x + \Delta x) - u(x))/\Delta x$) from one side to the other. The minus sign in (17) indicates
that heat will “flow” from regions of high temperature to regions of low temperature. In general, the thermal conductivity \(k\) (which is a measure of the material’s ability to transfer or “conduct” heat) may depend on \(t, x\) or even the temperature \(u\). Furthermore, the cross sectional area \(A\) may depend on \(x\) (a nonuniform geometry) and/or \(t\) (a “pulsating” solid or biological compartment). However, in our fundamental development of the heat equation to be presented in the next section, we shall assume that both \(k\) and \(A\) are constant (uniform in space and time).

### 13.3 Heat Equation

We begin by considering the unsteady heat transfer problem in one direction in a solid. To derive the conduction equation in one dimension, we refer to Figure 13, which depicts a small section of a one-dimensional cylindrical rod centered about an arbitrary point \(x\).

![Figure 13: Transient conduction in one-dimensional cylindrical rod.](image)

We make the following assumptions:

(i) Heat transfer is by conduction;

(ii) Heat transfer is along the \(x\)-axis;

(iii) Temperature is uniform over a cross-section;

(iv) We have perfect insulation, hence no heat is escaping from the sides of the cylindrical rod.

Let \(u(t, x)\) denote the temperature at \(x\) at a given time \(t\) and \(H\) denote the amount of heat (energy) in units of calories (a calorie is defined as the amount of energy required to raise 1 gm of water 1 °C). Heat may also be given in units of Joules (1 cal = 4.19 J or 1 J = 0.23885 cal). We expect the amount of heat in an element of mass to be proportional to both the mass and the temperature. This motivates the quantitative expression for heat:

\[
H = c_pmu, \quad (18)
\]
where $c_p$ is the specific heat, a constant of proportionality which depends on the material, and $m$ is the mass. The specific heat is given at a constant volume and has units $\frac{J}{kg \, K}$.

We are now ready to turn to energy balance in the small element of the volume between $x - \Delta x$ and $x + \Delta x$ as shown in Figure 13. Since the wall of the cylindrical rod is insulated, if we assume that there is no heat generated inside the cylinder, then we have

\[
\text{the net rate of heat accumulation} = \text{rate of heat input} - \text{rate of heat output}. \tag{19}
\]

We assume without loss of generality that heat flow is from left to right (i.e., $\frac{\partial u}{\partial x} < 0$).

The rate of heat input to the cylinder is

\[
\text{rate of heat input} = \dot{q}|_{x-\Delta x} = -kA \frac{\partial u}{\partial x}(t, x - \Delta x). \tag{20}
\]

Also,

\[
\text{rate of heat output} = \dot{q}|_{x+\Delta x} = -kA \frac{\partial u}{\partial x}(t, x + \Delta x). \tag{21}
\]

The rate of heat accumulation in the elemental volume $2A\Delta x$ is

\[
\text{rate of heat accumulation} = \frac{\partial H}{\partial t}, \tag{22}
\]

and by using the expression (18) for heat, we obtain

\[
\frac{\partial H}{\partial t} = \frac{\partial}{\partial t} (c_p m u) = \frac{\partial}{\partial t} (c_p (2\Delta x \rho A) u(t, x)) = 2\Delta x c_p \rho A \frac{\partial u(t, x)}{\partial t}, \tag{23}
\]

where $\rho$ denotes the mass density of the cylindrical rod. Substituting equations (20), (21), and (23) into (19) and dividing by $2\Delta x A$, we have

\[
\rho c_p \frac{\partial u}{\partial t} = \frac{k \frac{\partial u}{\partial x}(t, x + \Delta x) - k \frac{\partial u}{\partial x}(t, x - \Delta x)}{2\Delta x}.
\]

Letting $\Delta x \to 0$ we obtain

\[
\rho c_p \frac{\partial u}{\partial t} = \frac{\partial}{\partial x} (k \frac{\partial u}{\partial x}),
\]
or, since \( k \) is independent of \( x \),

\[
\frac{\partial u}{\partial t} = \left( \frac{k}{\rho c_p} \right) \frac{\partial^2 u}{\partial x^2}.
\]

This can be written as

\[
\frac{\partial u}{\partial t} = \alpha \frac{\partial^2 u}{\partial x^2}, \tag{24}
\]

where \( \alpha \equiv \frac{k}{\rho c_p} \) is the thermal diffusivity in \( \text{m}^2/\text{sec} \). Equation (24) is known as the one-dimensional heat equation. Since \( k \) is the material’s ability to conduct heat and \( \rho c_p \) is the volumetric heat capacity (ability of the material to store heat), the thermal diffusivity represents the ability of the material to conduct thermal energy relative to its ability to store it.

A similar derivation when the heat flow is from right to left (i.e., \( \frac{\partial u}{\partial x} > 0 \)) will give the same partial differential equation (24) for the heat conduction in a one-dimensional cylindrical rod.

Before turning to the three-dimensional version of the above quantitative description of heat transfer, we note that heat conduction and Fourier’s law are often discussed in terms of heat flux which is the rate of heat transfer (in the direction \( x \)) per unit cross sectional area and is given by

\[
\Phi = \frac{\dot{q}}{A} = -k \frac{\partial u}{\partial x}. \tag{25}
\]

For heat flux through a general (smooth) surface in three dimensions, the above formula (25) is generalized to have the form

\[
\Phi = -k \nabla u \cdot \hat{n}, \tag{26}
\]

where \( \hat{n} \) is the unit outward normal vector to the surface (\( \hat{n} = \pm \hat{i} \) in the above one-dimensional case).

We now consider a general region \( V \) and an arbitrary infinitesimal volume \( \Delta V \) enclosed by a surface \( \Delta S \) (see Figure 14).

We will formulate heat balance equations for the infinitesimal volume \( \Delta V \). First, we have

\[
\text{rate of heat accumulation in } \Delta V = \frac{\partial H}{\partial t} = \frac{\partial}{\partial t} \left( \int_{\Delta V} \rho c_p u \, dV \right) = \int_{\Delta V} \rho c_p \frac{\partial u}{\partial t} \, dV. \tag{27}
\]
Figure 14: (a) A general three-dimensional region. (b) An infinitesimal volume.

Also, from (26), the heat flux across the boundary \( \Delta S \) of \( \Delta V \) is given by

\[ \Phi = -k \nabla u \cdot \hat{n}. \]

If \( \nabla u \cdot \hat{n} \) is positive, we have heat flow into the infinitesimal element, so the rate of change of heat across a surface element \( dS \) is given by

\[ -\Phi dS = k \nabla u \cdot \hat{n} dS, \]

which is positive (i.e., the temperature \( u \) is increasing in the element along the \( \hat{n} \) direction.). In this case \( \Phi \) is negative, i.e., heat is entering the region. If \( \nabla u \cdot \hat{n} \) is negative, we have heat flow out of the element and the rate of change is again given by

\[ -\Phi dS = k \nabla u \cdot \hat{n} dS, \]

which is negative (i.e., the temperature \( u \) in the element is decreasing along the \( \hat{n} \) direction). In this case the flux \( \Phi \) is positive, i.e., heat is leaving the region. In either case, the rate of change of heat in the volume \( \Delta V \) is given by summing the rate (or the negative of the flux) across the boundary surface area:

\[
\frac{\partial H}{\partial t} = \int_{\Delta S} -\Phi dS = \int_{\Delta S} k \nabla u \cdot \hat{n} dS. \tag{28}
\]

By Gauss’ Theorem (the divergence theorem) \[98\] (see also Appendix B) we find that this last expression (28) can be rewritten as

\[
\frac{\partial H}{\partial t} = \int_{\Delta V} \nabla \cdot (k \nabla u) dV. \tag{29}
\]

Substituting equation (27) for the left side, we have

\[
\int_{\Delta V} \left[ \rho c_p \frac{\partial u}{\partial t} - \nabla \cdot (k \nabla u) \right] dV = 0
\]

35
for any arbitrary element \( \Delta V \) in \( V \). Since \( \Delta V \) is arbitrary, it follows that we must have

\[
\rho c_p \frac{\partial u}{\partial t} = \nabla \cdot (k \nabla u)
\]

in \( V \). For constant thermal conductivity \( k \), this equation is simplified to the following heat equation in three dimensions:

\[
\rho c_p \frac{\partial u}{\partial t} = k \nabla \cdot (\nabla u) = k \nabla^2 u,
\]

where \( \nabla^2 u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} \). For additional details on the development of the heat equation, the interested reader may consult [93] or [97].

13.4 Boundary conditions and initial conditions

Consider a simple ordinary differential equation (which is the steady-state case of equation (24))

\[
\frac{d^2 u}{dx^2} = 0.
\]

It has infinitely many solutions

\[
u(x) = c_1 x + c_2,
\]

where \( c_1 \) and \( c_2 \) are constants. Thus, to find a unique solution even in this simple case, one must impose additional conditions on the problem. To find a unique solution to the heat equation (24) or (30) we also must impose auxiliary equations. We choose these auxiliary equations to describe the state of our heat flow at time “zero” (the beginning of the experiment, for example) and the state of the flow on the boundary of our region. These equations are called initial conditions and boundary conditions, respectively.

We consider the one-dimensional heat equation (24) on a region \((0, L)\)

\[
\frac{\partial u}{\partial t} = \alpha \frac{\partial^2 u}{\partial x^2}, \quad 0 < x < L, \; t > 0.
\]

(31)

From equation (31), there is one time derivative, which implies that we need one initial condition specifying \( u \) for all \( x \) at a given time. In this case we do this for \( t = 0 \) and hence have

\[
u(0, x) = (x), \quad 0 < x < L.
\]

(32)
In addition, there are two spatial derivatives, which imply that we need two boundary conditions specifying $u$ for all $t$ at given values of $x$. For example, we might specify
\begin{equation}
    u(t, 0) = u_1(t) \text{ and } u(t, L) = u_2(t),
\end{equation}
and these are known as Dirichlet boundary conditions.

In many cases of interest, boundary conditions might be related to heat flux rather than to temperature. That is, we might specify fixed heat flux at one endpoint or both endpoints of the domain. For example, at $x = L$ we might impose
\begin{equation}
    \Phi(t, L) = -k \frac{\partial u}{\partial x}(t, L) = f(t).
\end{equation}

The condition (34) is called a Neumann boundary condition. A special case of (34) is
\begin{equation}
    k \frac{\partial u}{\partial x}(t, L) = 0,
\end{equation}
which means that the endpoint at $x = L$ is insulated.

Finally, we might combine both conditions (33) and (34) to obtain a condition of the form
\begin{equation}
    k \frac{\partial u}{\partial x}(t, L) + hu(t, L) = g(t),
\end{equation}
which is known as a Robin boundary condition. Here, the parameter $h$ has a physical meaning. It is well known that a hot piece of material will cool faster when air is blown or forced by the object. When the fluid or gas (air) outside the solid surface is forced or when we have natural convective flow, the rate of heat transfer from the solid to the fluid, or vice versa, is given by
\begin{equation}
    \dot{q} = hA(u_s - u_f),
\end{equation}
where $\dot{q}$ is the heat transfer rate in W, $A$ is the area in m$^2$, $u_s$ is the temperature of the solid surface in $\circ$K, $u_f$ is the average or bulk temperature of the fluid flowing by in $\circ$K and $h$ is the convective heat transfer coefficient or Newton cooling constant in W/m$^2$K.
The relation (36) is referred to as the "Newton’s law of cooling". Like other “laws”, it is not actually a law but one may think of it as a definition for $h$ based on empirical observations. Since we know that when a fluid flows by a solid surface, there is a thin film, which is almost stationary, adjacent to the solid wall which presents most of the resistance to heat transfer, the parameter $h$ is also often called the film coefficient or film conductance. In general, $h$ can not be predicted theoretically. It is a function of the system geometry, fluid properties, flow velocity, and, in some cases, the temperature difference. In Table 1 some values of $h$ are given for different mechanisms of heat transfer and materials (see for instance, [94, 95]).

If we next divide both sides of equation (36) by $A$, we obtain the convective heat flux

$$\Phi = \frac{\dot{q}}{A} = h(u_s - u_f).$$

So, we might specify the heat flux at one interface between the solid and fluid (or air) as

$$-\Phi(t, L) = h[u_f - u(t, L)],$$

which, after substitution of the form for heat flux (25), can be rewritten as

$$k \frac{\partial u}{\partial x}(t, L) + hu(t, L) = hu_f.$$

This equation is of the same form as the Robin boundary condition (35) and, hence, the meaning of the constant $h$ as discussed above. We note that, if $u_f > u(t, L)$, heat flows into the solid; otherwise, heat flows out of the solid (cooling).

Before we conclude this section, we will describe another type of boundary condition that occurs in some practical applications. This type of boundary condition is related to the third type of heat transfer mechanism - radiation heat transfer. We recall that this is basically an electromagnetic mechanism that allows energy to be transported with the speed of light through space. Since it consists of energy in the form of light waves, it obeys the same laws as does light. That is, it travels in straight lines and is
transmitted through vacuum and space. The associated quantitative law is given by the Stefan-Boltzmann law, expressed as

\[ \dot{q}_{\text{max}} = A \sigma u^4, \]  

(38)

where \( \dot{q}_{\text{max}} \) is the maximum rate of emitted heat in units W. The parameter \( \sigma \) is the Stefan-Boltzmann constant (= \( 5.676 \times 10^{-8} \) W/m\(^2\)K\(^4\)) and \( u \) is the (absolute) temperature (in °K) of the emitting surface. A body that achieves this rate in either emission or absorption is called a perfect radiator or black body, respectively. The actual emitted rate of a general surface is given by a somewhat smaller number

\[ \dot{q} = \epsilon A \sigma u^4 \]  

(39)

where \( 0 \leq \epsilon < 1 \), with \( \epsilon \) called the emissivity of the surface or body. When \( \epsilon < 1 \), we have a gray body or a gray surface. Bodies (surfaces) also absorb energy, for example, a black body (a perfect absorber) is defined as one that absorbs all radiant energy and reflects none. If \( \dot{q}_{\text{abs}}, \dot{q}_{\text{inc}} \) represent the rate of energy absorbed and rate of energy incident, then the surface absorptive property is characterized by a parameter \( \alpha \) called the absorptivity and is defined by

\[ \dot{q}_{\text{abs}} = \alpha \dot{q}_{\text{inc}}. \]

(Unfortunately, the same symbol \( \alpha \) often is used in the literature for both the conductive diffusivity and for the absorptivity.) For a gray surface defined by \( \alpha = \epsilon \), the net rate of heat exchange between a surface and its ambient gas is given by

\[ \dot{q}_{\text{net}} = \epsilon A \sigma (u_{\text{sur}}^4 - u_{\text{amb}}^4), \]

where \( u_{\text{sur}}, u_{\text{amb}} \) are the surface and ambient temperatures, respectively. Sometimes this net rate of transfer is written as

\[ \dot{q}_{\text{net}} = h_r A (u_{\text{sur}} - u_{\text{amb}}), \]

where the radiative heat transfer coefficient \( h_r \) is defined by

\[ h_r \equiv \epsilon \sigma (u_{\text{sur}} + u_{\text{amb}}) \left( u_{\text{sur}}^2 + u_{\text{amb}}^2 \right). \]

We note that treating \( h_r \) as a constant essentially linearizes the radiation rate equation. Finally, when radiation heat transfer occurs from the surface of a solid, it is usually accompanied by convective heat transfer unless the solid is in vacuum. The appropriate boundary condition is then given by

\[ k \frac{\partial u}{\partial n} \bigg|_{\partial \Omega} = h \left[ u_{\text{amb}} - u \bigg|_{\partial \Omega} \right] + \epsilon \sigma \left[ u_{\text{amb}}^4 - u^4 \bigg|_{\partial \Omega} \right]. \]
where $\partial \Omega$ is the closed surface enclosing the solid.

In practice, it is very difficult to ensure perfect insulation. In fact, in the design of the experiment, we have an uninsulated metal bar which is heated at one end and allows heat to escape along its entire length.

Under the assumption of no insulation, one can derive the following equation for the conduction of heat in the rod:

$$\rho c_p \frac{\partial u(t, x)}{\partial t} = k \frac{\partial^2 u(t, x)}{\partial x^2} - \frac{2(a + b)}{ab} h(u(t, x) - u_{\text{ambient}}), \quad 0 \leq x \leq L,$$

where $a$ and $b$ are the dimensions of the cross-sectional area of the rod. The minus term on the right hand side of the above equation comes from the heat loss term along the length of the rod, which should be modeled by the Newton’s law of cooling.
14 Statistically Based Model Comparison Techniques

In previous sections we have discussed techniques (e.g., residual plots) for investigating correctness of the assumed statistical model underlying the estimation (OLS or GLS) procedures used in inverse problems. To this point we have not discussed correctness issues related to choice of the mathematical model. However there are a number of ways in which questions related to the mathematical model may arise. In general, modeling studies [80, 81] can raise questions as to whether a mathematical model can be improved by more detail and/or further refinement. For example, one might ask whether one can improve the mathematical model by assuming more detail in a given mechanism (constant rate vs. time or spatially dependent rate – e.g., see [74] for questions related to time dependent mortality rates during sub-lethal damage in insect populations exposed to various levels of pesticides). Or one might question whether an additional mechanism in the model might produce a better fit to data–see [78, 79, 80] for diffusion alone or diffusion plus convection in cat brain transport in grey vs. white matter considerations. Before continuing, an important point must be made: In model comparison results outlined below, there are really two models being compared: the mathematical model and the statistical model. If one embeds the mathematical model in the wrong statistical model (for example, assuming constant variance when this really isn’t true), then the mathematical model comparison results using the techniques presented here will be invalid (i.e., worthless). An important remark in all this is that one must have the mathematical model one wants to simplify or improve (e.g., test whether $V = 0$ or not in the example below) embedded in the correct statistical model (determined in large part by the observation process), so that the comparison actually is only with regard to the mathematical model.

To provide specific motivation, we illustrate the formulation of hypothesis testing by considering a mathematical model for a diffusion-convection process. This model was proposed for use with experiments designed to study substance (labelled sucrose) transport in cat brains, which are heterogeneous, containing grey and white matter [80]. In general, the transport of substance in cat’s brains can be described by a PDE describing change in time and space. This convection/diffusion model, which is widely discussed in the applied mathematics and engineering literature, has the form

$$\frac{\partial u}{\partial t} + V \frac{\partial u}{\partial x} = D \frac{\partial^2 u}{\partial x^2}. \quad (40)$$

Here, the parameter $\vec{q} = (D, V)$, which belongs to some admissible parameter set $Q$, denotes the diffusion coefficient $D$ and the bulk velocity $V$ of the fluid, respectively.
Our problem: test whether the parameter $V$ plays a significant role in the mathematical model. That is, if the model (41) represents a diffusion-convection process, we seek to determine whether diffusion alone or diffusion plus convection best describes transport phenomena represented in cat brain data sets $\{y_{ij}\}$ for $\{u(t_i, x_j; \vec{q})\}$, the concentration of labelled sucrose at times $\{t_i\}$ and location $\{x_j\}$. We thus might wish to test the null hypothesis $H_0$ that diffusion alone best describes the data versus the alternative hypothesis $H_A$ that convection is also needed. We then may take $H_0 : V = 0$ and the alternative $H_A : V \neq 0$. Consequently, the restricted parameter set $Q_H \subset Q$ defined by

$$Q_H = \{\vec{q} \in Q : V = 0\}$$

will be important. To carry out these determinations, we will need some model comparison tests of analysis of variance (ANOVA) type [87] from statistics involving residual sum of squares (RSS) in least squares problems.

More generally, we the **diffusion-convection-reaction (DCR)** equation

$$\frac{\partial u}{\partial t} + V \frac{\partial u}{\partial x} = D \frac{\partial^2 u}{\partial x^2} + R. \quad (41)$$
15 HIV Models

16 Introduction

Human Immunodeficiency Virus (HIV) is a retrovirus that infects T-helper cells of the immune system and is the causative agent for Acquired Immune Deficiency Syndrome (AIDS). HIV and AIDS are among the world’s most serious public health concerns, affecting people of all demographics worldwide, with some regions impacted disproportionately. As of 2003, an estimated 38 million HIV-infected individuals are living worldwide, with approximately two-thirds in Africa, where 2.2 million people died from opportunistic infections related to AIDS in 2003 (UNAIDS 2004 Report on the Global HIV/AIDS Epidemic [31]). Despite many successful public health and clinical interventions since the first identification of HIV-positive patients in 1981, there remains no cure and the HIV/AIDS epidemic continues to grow.

Highly Active Antiretroviral Therapy (HAART), most commonly administered in the form of drug cocktails consisting of a protease inhibitor and at least one or more reverse transcriptase inhibitors, has been highly successful in suppressing HIV in many patients and therefore improving quality of life. However, contrary to dangerous popular myths, these drugs do not constitute a cure. While antiretroviral drugs are widely available in the United States and Western Europe, their cost and side effects may make their use challenging. In developing nations, UNAIDS estimates that only 7% of the infected population has access to HAART. Access to treatment for and education about this disease remain serious human rights issues around the world. Improved strategies are needed for efficient and appropriate use of drug therapy in both developed and underdeveloped countries.

Studies of the epidemiology of HIV and public health issues such as transmission (inter-host dynamics) are important. Equally important to investigate are the effective use and improvement of antiretroviral drugs, which depend on understanding viral behavior within each host, including pathways of infection and effects of drugs. Understanding intra-host viral and immune system pathways depends on knowledge from various biological areas including physiology and immunology. Mathematical models, when combined with statistically-based inverse problem techniques, can aid in quantifying dynamic physiologic and immunologic processes, correlating the scientific knowledge of these processes with observed patient behavior, and predicting patient outcomes. An example of such a modeling approach is given in this paper.

It is believed that the acute and early phases of HIV infection provide crucial information about immune responses and viral dynamics. In particular, long-term viral set points and speed of progression to AIDS may possibly be understood by studying these key periods. Motivated by clinical study data from patients observed during the crucial acute infection phase and beyond, we outline here a combined mathematical and statistical inverse problem approach for modeling HIV infection. We apply the methods to clinical
data and demonstrate the types of suggestions and conclusions one may draw from such an effort.

A number of patients for whom we have clinical data underwent therapy interruptions. Some of these drug holidays were unprescribed or single interruptions, while others were structured treatment interruptions (STIs) according to a study protocol. STI therapy protocols are currently being explored (not without controversy) as an alternative to continuous therapy with antiretrovirals since in addition to offering the benefit of reduced side effects, they may also serve to boost HIV-specific immune responses. We therefore incorporate STI protocols in our mathematical models. A good overview of the concept of STI and its applicability in various phases of HIV infection can be found in [44].

In previous work [29] we demonstrated that a differential equation model for in-host HIV infection dynamics can describe censored clinical data obtained from patients undergoing therapy interruptions. This entailed a process of parameter identification (estimation or model fitting) in order to determine values for the dynamic parameters in the model that will best describe the data. Model fitting in this manner yields valuable estimates of dynamic rates and quantities, for example the rate of growth of virus or infectivity contact rate, which might be used to differentiate between or explain patient behaviors. In this paper, we explore one of the most powerful features of mathematical models – the ability to assist in making predictions or understanding biological phenomena. We demonstrate how one could use longitudinal HIV viral load and CD4+ T-cell data gathered from a particular patient over a limited observation period, in conjunction with a biologically-based mathematical model, to make predictions about the patient’s long-term behavior. This might include the patient’s viral load or T-cell dynamics over time or a prediction of the long-term viral load set point.

In this way, our HIV model can be used to gain insight into potential clinical outcomes. For example, after calibration (i.e., parameter estimation), one could use the model to explore what would happen to a particular patient under various treatment strategies, including allowing the patient to remain completely off treatment.

17 Clinical Data Description

The data for our investigations come from a study of over 100 adults with symptomatic acute or early HIV-1 infection. These subjects were enrolled in a study based at Massachusetts General Hospital and associated regional centers and followed for varying lengths of time between 1996 and 2004. The study cohort is unique in that its members were all identified soon after initial infection, making its data particularly useful for understanding early viral dynamics and related immune responses. A principal goal of the clinical study is to assess the potential immunologic consequences of early treatment initiation, including preservation of HIV-specific CD4+ T-cells, extent of latent reservoir development, and homogeneity of viral population. Clinical researchers also strive to understand the role of early immune responses in long-term viral suppression.
Clinical and demographic data were collected at the time of study enrollment and blood draw assays of CD4+ T-lymphocyte count and RNA viral load performed at roughly monthly follow-up visits. Viral load was quantified with Reverse Transcriptase-Polymerase Chain Reaction (RT-PCR) methods using the commercially available HIV-1 Roche Amplicor or Chiron Quantiplex assay, yielding measurements in viral RNA copies per milliliter (ml). The standard assay has a linear range of 400 to 750,000 copies/ml, while the ultra-sensitive assay has a range of 50 to 100,000 copies/ml. The latter is typically employed when a measurement is below the 400 copies/ml limit of the standard assay, as is often the case for a patient successfully suppressing virus. Standard flow cytometry methods were employed to obtain total plasma CD4+ T-lymphocyte counts per microliter (μl) [40].

Nearly all subjects in the study underwent combination therapy with three or more antiretroviral drugs, although the precise regimen varied from patient to patient as dictated by the treating physician. Fourteen of the subjects underwent structured treatment interruptions according to a study protocol, including patients with identification numbers 2, 4, 5, 6, 10, 13, and 14 for whom immune responses were assessed during interruption [49]. Several others simply discontinued drugs at various points.

In this paper we focus on data from a subset of 45 patients wherein each patient has ten or more each of CD4 and viral load measurements in the first half of their longitudinal data. We denote this set by PS45 and summarize the data in Table 2, which includes the clinical identification number assigned to the patient, number of longitudinal viral load and CD4+ measurements, the total length of time from presentation to last observation, total number of days on and off treatment, and the number of periods (of any length) the patient was off and on therapy. The last two columns indicate the number of available data points during the first half of each patient’s time-series. The number of treatment interruptions varies drastically over the population and some patient records include an initial brief off-treatment phase after presentation before therapy commenced.

The treatment patterns and overall lengths of observation for each of the 45 patients are depicted in Figure 15. In these schematics, thicker lines denote on-treatment periods and the thinner lines, off-treatment.

We expect that some aspects of the mathematical model later considered are more readily validated in the context of treatment schemes with a balance between time on and time off treatment. Of the 45 patients considered in this paper, sixteen (those numbered 2, 4, 5, 6, 9, 10, 13, 14, 15, 23, 24, 26, 27, 33, 46, and 47) spend 30–70% time off treatment. Of these only patients 9, 15, and 47 do not spend appreciable time off treatment during the early half of their observation period.

Due to the linear range limits described above, the clinical viral load assays effectively have lower and upper limits of quantification. The upper limit is typically readily handled by repeatedly diluting the sample until the resulting viral load measurement is in range and then scaling. The lower limit, or left censoring point, however, directly influences the observed data. When a data point is left-censored (below the lower limit of
Table 2: Summary of data for 45 patients with ten or more each CD4 and viral load measurements in the first half of their longitudinal data, ordered by clinical identification number. Includes number of measurements, duration of observation, time on versus off treatment, and the number of measurements in the early half of the time series.

<table>
<thead>
<tr>
<th>pat num</th>
<th>num meas</th>
<th>total days</th>
<th>days on/off</th>
<th>periods on/off</th>
<th># in 1st half</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>VL CD4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>102/84</td>
<td>1527</td>
<td>1316/211</td>
<td>4/3</td>
<td>49/41</td>
</tr>
<tr>
<td>2</td>
<td>107/82</td>
<td>1966</td>
<td>902/1064</td>
<td>2/2</td>
<td>67/48</td>
</tr>
<tr>
<td>3</td>
<td>76/61</td>
<td>1943</td>
<td>1589/354</td>
<td>3/2</td>
<td>40/31</td>
</tr>
<tr>
<td>4</td>
<td>154/107</td>
<td>1919</td>
<td>1248/671</td>
<td>4/4</td>
<td>81/58</td>
</tr>
<tr>
<td>5</td>
<td>158/115</td>
<td>2061</td>
<td>1067/994</td>
<td>4/4</td>
<td>88/66</td>
</tr>
<tr>
<td>6</td>
<td>143/111</td>
<td>1839</td>
<td>923/916</td>
<td>4/5</td>
<td>85/64</td>
</tr>
<tr>
<td>7</td>
<td>23/22</td>
<td>1932</td>
<td>1924/8</td>
<td>1/1</td>
<td>15/14</td>
</tr>
<tr>
<td>8</td>
<td>34/33</td>
<td>1672</td>
<td>1668/4</td>
<td>1/1</td>
<td>24/23</td>
</tr>
<tr>
<td>9</td>
<td>32/32</td>
<td>1626</td>
<td>1112/514</td>
<td>2/3</td>
<td>24/24</td>
</tr>
<tr>
<td>10</td>
<td>73/63</td>
<td>1711</td>
<td>582/1129</td>
<td>1/1</td>
<td>57/47</td>
</tr>
<tr>
<td>12</td>
<td>24/19</td>
<td>1575</td>
<td>1540/35</td>
<td>2/1</td>
<td>18/15</td>
</tr>
<tr>
<td>13</td>
<td>64/55</td>
<td>914</td>
<td>537/377</td>
<td>3/3</td>
<td>34/28</td>
</tr>
<tr>
<td>14</td>
<td>136/91</td>
<td>1637</td>
<td>659/978</td>
<td>3/3</td>
<td>78/57</td>
</tr>
<tr>
<td>15</td>
<td>46/46</td>
<td>1659</td>
<td>932/727</td>
<td>1/1</td>
<td>23/23</td>
</tr>
<tr>
<td>18</td>
<td>32/30</td>
<td>1545</td>
<td>1545/0</td>
<td>1/0</td>
<td>26/25</td>
</tr>
<tr>
<td>19</td>
<td>21/19</td>
<td>1430</td>
<td>1416/14</td>
<td>1/1</td>
<td>19/17</td>
</tr>
<tr>
<td>20</td>
<td>29/27</td>
<td>1581</td>
<td>1469/112</td>
<td>1/2</td>
<td>18/17</td>
</tr>
<tr>
<td>21</td>
<td>38/36</td>
<td>1433</td>
<td>1412/21</td>
<td>1/1</td>
<td>26/24</td>
</tr>
<tr>
<td>23</td>
<td>37/36</td>
<td>1605</td>
<td>671/834</td>
<td>4/5</td>
<td>21/21</td>
</tr>
<tr>
<td>24</td>
<td>36/35</td>
<td>1436</td>
<td>841/595</td>
<td>4/3</td>
<td>27/26</td>
</tr>
<tr>
<td>25</td>
<td>83/60</td>
<td>1412</td>
<td>1255/157</td>
<td>4/4</td>
<td>46/33</td>
</tr>
<tr>
<td>26</td>
<td>100/72</td>
<td>1434</td>
<td>754/680</td>
<td>3/4</td>
<td>59/37</td>
</tr>
<tr>
<td>27</td>
<td>36/35</td>
<td>1379</td>
<td>591/788</td>
<td>2/2</td>
<td>21/21</td>
</tr>
<tr>
<td>29</td>
<td>34/34</td>
<td>1024</td>
<td>1017/7</td>
<td>1/1</td>
<td>25/25</td>
</tr>
<tr>
<td>30</td>
<td>16/13</td>
<td>841</td>
<td>837/4</td>
<td>1/1</td>
<td>12/10</td>
</tr>
<tr>
<td>31</td>
<td>30/30</td>
<td>1256</td>
<td>1228/28</td>
<td>2/2</td>
<td>16/16</td>
</tr>
<tr>
<td>32</td>
<td>33/33</td>
<td>1230</td>
<td>1209/21</td>
<td>1/1</td>
<td>21/20</td>
</tr>
<tr>
<td>33</td>
<td>75/52</td>
<td>1302</td>
<td>658/644</td>
<td>4/4</td>
<td>54/34</td>
</tr>
<tr>
<td>34</td>
<td>24/23</td>
<td>1174</td>
<td>1173/1</td>
<td>1/1</td>
<td>14/13</td>
</tr>
<tr>
<td>36</td>
<td>33/31</td>
<td>1167</td>
<td>1161/6</td>
<td>1/1</td>
<td>21/19</td>
</tr>
<tr>
<td>37</td>
<td>25/25</td>
<td>1146</td>
<td>1139/7</td>
<td>1/1</td>
<td>18/18</td>
</tr>
<tr>
<td>39</td>
<td>29/28</td>
<td>1023</td>
<td>910/113</td>
<td>3/3</td>
<td>20/20</td>
</tr>
<tr>
<td>41</td>
<td>22/21</td>
<td>717</td>
<td>688/29</td>
<td>2/1</td>
<td>16/15</td>
</tr>
<tr>
<td>42</td>
<td>30/30</td>
<td>1218</td>
<td>1170/48</td>
<td>2/1</td>
<td>17/17</td>
</tr>
<tr>
<td>43</td>
<td>28/29</td>
<td>1134</td>
<td>1060/74</td>
<td>1/1</td>
<td>15/16</td>
</tr>
<tr>
<td>45</td>
<td>46/28</td>
<td>499</td>
<td>418/81</td>
<td>2/2</td>
<td>14/12</td>
</tr>
<tr>
<td>46</td>
<td>100/55</td>
<td>1004</td>
<td>496/508</td>
<td>3/3</td>
<td>39/24</td>
</tr>
<tr>
<td>47</td>
<td>23/23</td>
<td>1002</td>
<td>496/506</td>
<td>1/2</td>
<td>10/10</td>
</tr>
<tr>
<td>52</td>
<td>20/19</td>
<td>708</td>
<td>674/34</td>
<td>1/1</td>
<td>13/12</td>
</tr>
<tr>
<td>54</td>
<td>25/25</td>
<td>878</td>
<td>868/10</td>
<td>1/1</td>
<td>15/14</td>
</tr>
<tr>
<td>55</td>
<td>14/14</td>
<td>806</td>
<td>748/58</td>
<td>1/1</td>
<td>10/10</td>
</tr>
<tr>
<td>60</td>
<td>19/18</td>
<td>746</td>
<td>720/26</td>
<td>1/1</td>
<td>11/10</td>
</tr>
<tr>
<td>65</td>
<td>18/17</td>
<td>755</td>
<td>728/27</td>
<td>1/1</td>
<td>13/12</td>
</tr>
<tr>
<td>75</td>
<td>16/15</td>
<td>549</td>
<td>549/28</td>
<td>3/3</td>
<td>11/10</td>
</tr>
<tr>
<td>84</td>
<td>16/15</td>
<td>461</td>
<td>461/0</td>
<td>1/0</td>
<td>11/10</td>
</tr>
</tbody>
</table>
Figure 15: Treatment protocols and observation periods for patients in PS45. Thick (green) lines denote on-treatment periods whereas thin (red) lines denote off-treatment.
quantification), the only available knowledge is that the true measurement is between zero and the limit of quantification $L_*$ for the assay. Those at hand have two limits of quantification, $L_1 = 400$ copies/ml for the standard and $L_2 = 50$ copies/ml for the ultra-sensitive assay. These are illustrated in sample data from patient 6 shown in Figure 16, where censored data points are those appearing identically on the horizontal censoring lines $L_1 = 400, L_2 = 50$. A statistical methodology for handling this type of censored data is described below in Section 18.2.

The observation times and intervals vary substantially between patients. The sample data in Figure 16 also reveal that observations of viral load and CD4 may not have been made at the same time points, so in general for patient number $j$ we have CD4$^+$ T-cell data pairs $(t_{1j}^{ij}, y_{1j}^{ij}), i = 1, \ldots, N_j^1$ and (potentially at different times) viral RNA data pairs $(t_{2j}^{ij}, y_{2j}^{ij}), i = 1, \ldots, N_j^2$. 

Figure 16: Patient 6 CD4$^+$ T-cell and viral load data, including censor points (lines at $L_1 = 400, L_2 = 50$) for viral load, and periods of on-therapy (solid lines on axis) and periods of off-therapy (dashed line on axis).
18 HIV Model and Inverse Problem Techniques

18.1 Model description

Many HIV models have been considered in the literature, including those surveyed in [35] and [48]. To demonstrate the potential predictive ability of such mathematical models, we employ the model developed in [26], subsequently modified in [29], and depicted in Figure 17; other models could be readily treated in our framework. The model compartments are denoted by variables \( T_1 \) (type 1 target cells, e.g., CD4\(^+\) T-cells, cells/\( \mu l \)), \( T_2 \) (type 2 target cells, e.g., macrophages, cells/\( \mu l \)), \( V_I \) (infectious free virus, RNA copies/ml), \( V_{NI} \) (non-infectious free virus, RNA copies/ml), and \( E \) (cytotoxic T-lymphocytes, cells/\( \mu l \)). A superscript asterisk (*) denotes infected cells. The available clinical data include total CD4\(^+\) T-cell count, represented by the sum \( T_1 + T_1^* \), and total free virus, \( V_I + V_{NI} \).

![Figure 17: Schematic of compartmental HIV infection dynamics model. Only key pathways are indicated in the schematic — for further details, see the system of differential equations (42) below.](image)

While the remaining compartments \( T_2, T_2^*, \) and \( E \) were not observed in the data used in this paper, they are important for modeling and predicting long-term longitudinal data. The presence of a secondary target cell population \( T_2 \) helps to satisfy a modeling requirement suggested by Callaway and Perelson [35] in their 2002 review paper: a reasonable model of HIV infection predicts a non-zero steady-state viral load, even in
the presence of effective drug therapy. Patients subjected to drug therapy often successfully suppress virus for a long time, potentially at undetectable levels. However, some reservoir or mechanism exists that almost invariably causes the virus to grow out to detectable levels upon removal of drug therapy. Hence one does not expect incorporation of drug therapy in the model, at a sensible efficacy, to drive the viral load to zero, but rather reduce it considerably, perhaps below the assay limits of quantification. One way to incorporate this is shown in Figure 17, where there are two co-circulating populations of target cells, potentially representing CD4+ T-lymphocytes ($T_1$) and macrophages or other HIV-targeted cells ($T_2$). The two cell populations may have different activation requirements or susceptibility to drug therapy, represented by the different rate constants, thus potentially creating a non-zero, but low viral load steady state. This is crucial for modeling our long time horizon data, where patients may remain on treatment for an extended time. The differential efficacy also enables the model to exhibit reasonable sensitivity of the viral load equilibrium to treatment efficacy. For a survey of models and discussion of which exhibit reasonable sensitivity to drug efficacy, consult [35].

The documented importance of the immune system in responding to HIV infection (and especially its apparent crucial role during structured treatment interruptions) strongly motivates the inclusion of at least one model compartment representing immune response to the pathogen. We therefore include a measure $E$ of cytotoxic T-lymphocyte (CTL) CD8+ response to HIV infection. While the presently available data do not directly quantify the presence of HIV-specific CTLs, these immune responders are important for control of infected cells and may eventually be correlated to available epitope-challenge data. It is known that the immune response system is much more complicated than as represented in a single (composite) compartment denoted as CTL effectors $E$. Indeed, while present knowledge is incomplete, there are strong indications that a more complex modeling view of immune response involving naive and activated classes of CD4+ and HIV-specific CD8+ cells as well as memory and latent reservoir classes will be important in understanding the chronic versus acute response of the immune system to HIV-1 infection [43, 46].

The corresponding compartmental ordinary differential equation (ODE) model for in-host HIV infection dynamics is given by (42). This model is essentially one suggested in Callaway–Perelson [35], but includes an immune response compartment and dynamics as suggested by Bonhoeffer, et. al. [34]. This compartment, denoted by $E$, represents
The adapted system of ODEs is given by

\[
\begin{align*}
\dot{T}_1 &= \lambda_1 - d_1 T_1 - (1 - \bar{\epsilon}_1(t)) k_1 V_I T_1 \\
\dot{T}_2 &= \lambda_2 - d_2 T_2 - (1 - f \bar{\epsilon}_1(t)) k_2 V_I T_2 \\
\dot{T}_1^* &= (1 - \bar{\epsilon}_1(t)) k_1 V_I T_1 - \delta T_1^* - m_1 E T_1^* \\
\dot{T}_2^* &= (1 - f \bar{\epsilon}_1(t)) k_2 V_I T_2 - \delta T_2^* - m_2 E T_2^* \\
\dot{V}_I &= (1 - \bar{\epsilon}_2(t)) 10^3 N_T \delta (T_1^* + T_2^*) - c V_I \\
&\quad - (1 - \bar{\epsilon}_1(t)) 10^3 k_1 T_1 V_I - (1 - f \bar{\epsilon}_1(t)) 10^3 k_2 T_2 V_I \\
\dot{V}_{NI} &= \bar{\epsilon}_2(t) 10^3 N_T \delta (T_1^* + T_2^*) - c V_{NI} \\
\dot{E} &= \lambda_E + \frac{b_E (T_1^* + T_2^*)}{(T_1^* + T_2^*) + K_b} E - \frac{d_E (T_1^* + T_2^*)}{(T_1^* + T_2^*) + K_d} E - \delta_E E,
\end{align*}
\]

(42a)  
(42b)  
(42c)  
(42d)  
(42e)  
(42f)  
(42g)

together with an initial condition vector

\[(T_1(0), T_1^*(0), T_2(0), T_2^*(0), V_I(0), V_{NI}(0), E(0))^T.\]

Here the factors $10^3$ are introduced to convert between microliter and milliliter scales, preserving the units from some of the published papers.

As is common in models of HIV infection, infected cells $T_i^*$ result from encounters between uninfected target cells $T_i$ and infectious free virus $V_I$ in a well-mixed environment. As noted above, this model involves two co-circulating populations of target cells, perhaps representing CD4$^{+}$ T-lymphocytes ($T_1$) and macrophages ($T_2$). The natural infection rate $k_i$ may differ between the two populations, which could account for suspected differences in activation rates between lymphocytes and macrophages. The treatment factor $\bar{\epsilon}_1(t)$, described further below, represents a reverse transcriptase inhibitor (RTI) that blocks new infections and is potentially more effective in population 1 ($T_1^*$) than in population 2 ($T_2^*$), where the efficacy is $f \bar{\epsilon}_1$, with $f \in [0, 1]$. The differences in infection rates and treatment efficacy help create a low, but non-zero, infected cell steady state for $T_2^*$, which is commensurate with the idea that macrophages may be an important source of virus after T-cell depletion. The populations of uninfected target cells $T_1$ and $T_2$ may have different source rates $\lambda_i$ and natural death rates $d_i$.

Free virus particles are produced by both types of infected cells, which we assume produce virus at the same rate (again this could be readily generalized to account for
different productivity). In this model, virus may leave the $V_I$ compartment due to natural death at rate $c$ or via infecting a target cell (at rate $k_i T_i$). The action of a protease inhibitor (PI), which causes infected cells to produce non-infectious virus $V_{NI}$ is modeled by $\tilde{c}_2$. Tracking non-infectious virus is important because the clinically-measured viral load data for patients includes total free virus (sum of infectious $V_I$ and non-infectious $V_{NI}$).

Finally, the immune effectors $E$ (CTLs), are produced in response to the presence of infected cells and existing immune effectors. The immune response assumed here is similar to that suggested by Bonhoeffer, et al., in their 2000 paper [34], with a Michaelis-Menten type saturation nonlinearity. (Such a saturation type nonlinearity might be more biologically realistic in place of the product nonlinearities used elsewhere in the model, but to date our computations do not suggest a need for them at the present levels of modeling.) The infected cell-dependent death term in the immune response represents immune system impairment “at high virus load”. In [34] the authors present simulations which suggest that a model with this immune reponse structure and a latently infected cell compartment can exhibit transfer between “healthy” and “unhealthy” locally stable steady states via STI, making it a good candidate for our investigation. (Indeed, further investigations [28, 33] with (42) substantiate that active control through optimal or suboptimal STI therapies can readily effect such a transfer.) We add a source term $\lambda_E$ to create a non-zero off-treatment steady state for $E$, rather than explicitly modeling immune memory. While immune effectors are not inherently present in the absence of pathogen, they persist at low levels during infection. We note that other immune response models, such as those considered by Wodarz-Nowak [51] or Nowak-Bangham [47] could be substituted if desired. However, the latter does not appear to admit multiple stable off-treatment steady states.

The immune response we model is that of cytotoxic T-lymphocytes. CTLs act by lysing infected cells, causing them to explode. Thus they remove infected cells from the system in the equations for $\dot{T}_1$ and $\dot{T}_2$, at rates $m_1$ and $m_2$, respectively. Unlike interferons, they do not directly target free virus, so there is no interaction term with the virus compartment.

In this dynamical system, the treatment factors $\bar{c}_1(t) = \epsilon_1 u(t)$ and $\bar{c}_2(t) = \epsilon_2 u(t)$ represent the effective treatment impact, consisting of efficacy factors $\epsilon_1, \epsilon_2$ and a time-dependent treatment function $0 \leq u(t) \leq 1$ representing HAART drug level, where $u(t) = 0$ is fully off and $u(t) = 1$, fully on. Figure 18 depicts a sample time-varying treatment protocol representing structured therapy interruption. The relative effectiveness of RTIs is modeled by $\epsilon_1$ and that of PIs by $\epsilon_2$. Since HIV treatment is nearly always administered as combination therapy, we do not consider the possibility of monotherapy, even for a limited period of time, though this could be implemented by considering separate treatment functions $u_1(t), u_2(t)$. In the case of model fitting, the treatment protocol $u(t)$ is dictated by the clinical records for each patient. For a more thorough description of this model and its generalizations, the interested reader is referred to [26, 29].
Figure 18: Sample control input (treatment protocol) $u(t)$ representing structured treatment interruption. This is a schematic in that interruption periods need not be periodic and one might assume more smooth ramp functions for the absorption and dissipation of the drug.

In this paper, $\bar{x}$ will denote the vector of solutions to the ODE system (42); that is,

$$\bar{x}(t) = (T_1(t), T_1^*(t), T_2(t), T_2^*(t), V_I(t), V_{NI}(t), E(t))^T,$$

where components 1–4 of $\bar{x}$ are on a cells/$\mu$L scale, 5 and 6 (corresponding to $V_I$ and $V_{NI}$) on a copies/ml scale, and 7 on a cells/$\mu$L scale. The differential equation model (42) can therefore be represented by

$$\frac{d\bar{x}}{dt} = \bar{g}(t, \bar{x}; q),$$

with $q$ denoting model dynamic parameters and $\bar{g}$ the vector of derivatives. Model fits will be to the base-10 logarithm scale of these quantities ($x = \log_{10}\bar{x}$) and in general, variables with an overbar will denote an unscaled quantity and those without denote $\log_{10}$-transformed or scaled variables.

### 18.2 Inverse problem and censored data techniques

We wish to use the HIV model (42) to describe clinical data and make predictions, but it must first be “calibrated” to patient data by estimating appropriate parameters. That is, we use individual patient data (either partial or full longitudinal sets) to carry out inverse or parameter estimation problems to obtain patient-specific parameter estimates in the model. In this section we describe the standard nonlinear least squares method for doing so and then modify it with a method for treating the censored data measurements. As noted in the data description, in performing an inverse problem we do not have the luxury of observing the full vector of model states at each measurement time. Given an observation operator appropriate for the data, let $\bar{x}$ denote native model solutions and $\bar{z} = O\bar{x}$ be the observed model solutions. Recall that the number of observations might vary from patient to patient, so for each of $N_P$ patients, we have for $j = 1, \ldots, N_P$, times
\( \{ t_{ij}^1, i = 1, \ldots N_j^1 \} \) for CD4 measurements \( y_1(t) \) and \( \{ t_{ij}^2, i = 1, \ldots N_j^2 \} \) for viral RNA measurements \( y_2(t) \). We fit the model using the base-10 logarithm of these quantities: 
\[
x = \log_{10} x, \quad z = \log_{10} z, \quad y = \log_{10} y.
\]
Note that throughout our discussions we use subscripts 1 and 2 for parameters, data or indices to distinguish CD4 and viral RNA quantities, respectively. Superscripts \( j \) will be used to denote corresponding patient quantities for the \( j^{th} \) patient.

The inverse problem method will employ data from a single patient \( j \) in order to estimate one or more parameters \( (q) \). In this case, for each fixed patient \( j \), the goal is to fit the ODE model to his/her data by minimizing the cost criterion

\[
q^* j = \arg \min_{q \in Q} J(q) = \sum_{s=1}^2 \frac{1}{N_j^s} \sum_{i=1}^{N_j^s} \left| z_s(t_{ij}^s; q) - y_{ij}^s \right|^2
\]

over an admissible parameter set \( Q \subset \mathbb{R}^p \) to obtain optimal estimates for the \( p \)-vector parameters \( q \). This is the typical nonlinear least squares formulation, where \( J(q) \) depends through \( z \) on the solution to the nonlinear system of differential equations. There is a substantial literature on such problems [32], and in particular, many sampling- and gradient-based methods are available to iteratively solve (44) for \( q^* j \) (see Kelley [41] and the references therein). From a statistical point of view, minimizing (44) corresponds to maximum likelihood estimation of \( q \) assuming that the log-scaled measurements \( y_{ij}^s \) are normally distributed, i.e.,

\[
y_{ij}^s \sim N(z_s(t^i; q^0), \sigma_s^2), \quad s = 1, 2,
\]

for some the true underlying parameter values \( q^0 \) and variance \( \sigma_s^2 \), where (i) \( \sigma_1^2 = \sigma_2^2 \), and both log transformed CD4 and viral RNA measurements are (ii) independent across time (so not serially correlated) and (iii) independent of each other at each time. Assumptions (ii) and (iii) are reasonable approximations if it is assumed that the dominant source of variation in \( y_s \) values about \( z_s \) is assay error, with joint behavior at each time and across time of error-free CD4 and viral RNA values dictated primarily by the model. Assumption (i) is likely violated, as variation in CD4 and viral RNA measurements due to assay error is apt to be different. To take this into account would involve weighting each summand indexed by \( s \) in \( q^* j \) by estimates of \( 1/\sigma_s^2 \) and require some modification to the algorithm described next; for demonstration purposes, we focus on (44), recognizing that failure to weight in this manner may result in estimators for \( q \) that are less precise.

When viral load measurements are below the limit of quantification for the assay used, the observed values \( y_{ij}^2 \) do not represent the true data value and come with knowledge of censoring included. We must therefore modify the optimization problem to include this information. This may be accomplished by employing standard methods for censored data regression analysis [30, 50] as follows.

Unscaled measurements of viral load (second observed component, \( y_2^2 \)) are censored when below the limit of quantification, at either \( L1 = 400 \) or \( L2 = 50 \). In handling the
censored data, we exploit the assumption that the log-scaled measurements are normally distributed. Denote the log-scaled censoring points by \( L_1 = \log_{10} \bar{L}_1 \), \( L_2 = \log_{10} \bar{L}_2 \). For censored data points, the available knowledge is that the observed value \( y_i^2 \leq L_i \), where \( L_i \) denotes the relevant censoring point (\( L_i = L_1 \) or \( L_i = L_2 \)) at time \( t_i \).

In this context we observe pairs \((w_i, \chi_i)\), \( i = 1, \ldots, N \), where

\[
  w_i = \begin{cases} 
  y_i^2 & \text{if } y_i^2 > L_i \\
  L_i & \text{if } y_i^2 \leq L_i 
  \end{cases}
\]

\[
  \chi_i = I\{y_i^2 > L_i\},
\]

and \( I_A \) is the indicator function for the set \( A \). Defining the standard normal pdf \( \phi(\xi) = \frac{1}{\sqrt{2\pi}} e^{-\xi^2/2} \) with corresponding cdf \( \Phi(\xi) = \int_{-\infty}^{\xi} \phi(s)ds \), we have that the viral load portion of the likelihood function for \((q, \sigma_2)\) given the observations \( w_i \) is (see Sections 2.2.3 and 3.6 of [39] or Chapter 12 of [42])

\[
  \tilde{\mathcal{L}}(q, \sigma_2) = \prod_{i=1}^{N} \left[ \frac{1}{\sigma_2} \phi \left( \frac{w_i - z_i^2}{\sigma_2} \right) \right]^\chi_i \left[ \Phi \left( \frac{w_i - z_i^2}{\sigma_2} \right) \right]^{1-\chi_i},
\]

where the first term accounts for the probability of observing \( w_i \) given that it is un-censored and the second term the probability that the observation is in the interval \((-\infty, L_i)\) when censored. This is using a truncated normal distribution for the censored measurements. The log-likelihood is

\[
  \mathcal{L}(q, \sigma_2) = \sum_{i=1}^{N} \left[ \chi_i \left\{ \log \phi \left( \frac{w_i - z_i^2}{\sigma_2} \right) - \log \sigma_2 \right\} + (1 - \chi_i) \left\{ \log \Phi \left( \frac{w_i - z_i^2}{\sigma_2} \right) \right\} \right]
\]

\[
  = \sum_{i=1}^{N} \left[ \chi_i \left\{ \log \phi \left( \frac{y_i^2 - z_i^2}{\sigma_2} \right) - \log \sigma_2 \right\} + (1 - \chi_i) \left\{ \log \Phi \left( \frac{L_i - z_i^2}{\sigma_2} \right) \right\} \right]
\]

(45)

which we maximize to estimate \( q \) and \( \sigma_2 \). This is analogous to the typical log likelihood estimator in the absence of a limit of detection, where

\[
  \mathcal{L}(q, \sigma_2) = -\frac{N}{2} \log 2\pi - \frac{N}{2} \log \sigma_2 - \sum_{i=1}^{N} \frac{(y_i^2 - z_i^2)^2}{2\sigma_2^2}.
\]

(46)

However, while maximizing (46) in the parameters \( q \) is equivalent to minimizing the sum of squared residuals \( \sum_{i=1}^{N} (y_i^2 - z_i^2)^2 \) (typical least squares such as (44)) and the estimation of \( q \) and \( \sigma_2 \) decouple, maximizing (45) is not as simple, since a joint estimation of \( q \) and \( \sigma_2 \) must be performed. Maximizing (45) is possible with the Expectation Maximization (EM) algorithm [36, 45], which iteratively updates the estimates of \( q \) and \( \sigma_2 \) until the maximum is achieved.
Step 2

First, with the assumptions about distributions made above, let \( \xi^i = \frac{L - z_i^2}{\sigma_2} \) and \( \Lambda(\xi^i) = \frac{\phi(\xi^i)}{\Phi(\xi^i)} \), and use properties of a truncated normal distribution to obtain

\[
E \left[ y_i^2 | y_i^2 \leq L \right] = z_i^2 - \sigma_2 \Lambda(\xi^i), \text{ and} \\
E \left[ (y_i^2)^2 | y_i^2 \leq L \right] = (z_i^2)^2 - 2\sigma_2 z_i^2 \Lambda(\xi^i) - \sigma_2^2 \xi^i \Lambda(\xi^i) + \sigma_2^2.
\]

These can be used to obtain updates \((\hat{y}^i, \hat{r}^i)\) of the data points and estimates of squared residuals for the second observed state given by

\[
\hat{y}^i = \chi^i y_i^2 + (1 - \chi^i)E \left[ y_i^2 | y_i^2 \leq L \right] = \chi^i y_i^2 + (1 - \chi^i) \left[ z_i^2 - \sigma_2 \Lambda(\xi^i) \right] 
\]

and

\[
\hat{r}^i = \chi^i E \left[ (y_i^2 - z_i^2)^2 \right] + (1 - \chi^i)E \left[ (y_i^2 - z_i^2)^2 | y_i^2 \leq L \right] = \chi^i (y_i^2 - z_i^2)^2 + (1 - \chi^i) \left\{ E \left[ (y_i^2)^2 | y_i^2 \leq L \right] - 2z_i^2 E \left[ (y_i^2) | y_i^2 \leq L \right] + (z_i^2)^2 \right\} \\
= \chi^i (y_i^2 - z_i^2)^2 + (1 - \chi^i)\sigma_2^2 \left[ 1 - \xi^i \Lambda(\xi^i) \right].
\]

We can thus outline the EM Algorithm as follows.

**Algorithm 18.1** Expectation Maximization (EM) Algorithm

**Step 1** (Initialize) Create adjusted data \( \tilde{y}^i \) by replacing censored \( y_i^2 \) values (those for which \( \chi^i = 0 \)) by \( L^i/2 \), and use ordinary least squares to estimate \( \hat{q}^{(0)} \) using both CD4 data \( y_i^2, i = 1, \ldots, N_1 \), and viral RNA data \( \tilde{y}^i, i = 1, \ldots, N_2 \), (which includes replaced censored values). Obtain an initial estimate for \( \sigma_2^2 \) from

\[
(\hat{\sigma}_2^{(0)})^2 = \frac{1}{N_2} \sum_{i=1}^{N_2} \left| \tilde{y}^i - z_2(t_2^i; \hat{q}^{(0)}) \right|^2.
\]

Set \( k = 0 \).

**Step 2** Define \( \tilde{z}_2^{(k)} = z_2(t^i, \hat{q}^{(k)}) \) and \( \hat{\xi}^{(k)} = \frac{L - \tilde{z}_2^{(k)}}{\hat{\sigma}_2^{(k)}} \), and update the data and residuals by

\[
\tilde{y}^{(k)} = \chi^i y_i^2 + (1 - \chi^i) \left[ \tilde{z}_2^{(k)} - \hat{\sigma}_2^{(k)} \Lambda(\hat{\xi}^{(k)}) \right] \\
\tilde{r}^{(k)} = \chi^i (y_i^2 - \tilde{z}_2^{(k)})^2 + (1 - \chi^i) \left[ \hat{\sigma}_2^{(k)} \right]^2 \left( 1 - \hat{\xi}^{(k)} \Lambda(\hat{\xi}^{(k)}) \right).
\]

1. Update the estimates to \( \hat{q}^{(k+1)}, \hat{\sigma}_2^{(k+1)} \) by performing ordinary least squares minimization in the parameters \( q \)

\[
\hat{q}^{(k+1)} = \arg \min_q \sum_{i=1}^{N_1} |y_i^1 - z_1(t_1^i; q)|^2 + \sum_{i=1}^{N_2} |\tilde{y}^{(k)} - z_2(t_2^i; q)|^2
\]

56
and computing

$$(\sigma_2^{(k+1)})^2 = \frac{1}{N_2} \sum_{i=1}^{N_2} \tilde{r}_i^{(k)}.$$ 

If relative changes in $\hat{q}$ and $\hat{\sigma}$ are small, terminate. Otherwise set $k = k + 1$ and then go to Step 2.

This iterative process yields estimates of the parameters, variance, and expected values of the data at times where censored observations were recorded. This information can then be used to compute standard errors and confidence intervals on parameter estimates. While this is not pursued here, see [29] for further discussions and results.

19 Model fits and sample predictions

Fitting patient data using half the longitudinal data and then extrapolating over the whole time horizon often yields results similar to fitting the entire data set, supporting the model’s predictive ability. In Figures 19 and 20 corresponding to patients 14 and 4, respectively, we see model fits obtained using half and full time series data. Both these patients undergo two treatment interruptions during the early half of their data. The fit to viral load data is nearly the same regardless of whether half or all of the data are used. The fits to T-cell data are qualitatively different, but using half or all the data both yield plausible fits to data. For patient 14, the predicted viral load during the final off-treatment phase is within 1 log of the observed data. Similar results are obtained for other patients undergoing multiple interruptions, e.g., see figures for patients 2 and 6 in the appendix of the technical report CRSC-TR05-40 [27] where fits for each of the patients in our data set are given.

While calibrated solely with total virus and total T-cell count data, the model also suggests dynamics for the other (unobserved compartments). For parameters estimated using half of the data from patient 4, Figure 21 presents the model dynamics for target cell population 1, target cell population 2, total virus, and immune response $E$.

From Figure 22, observe that even with two interruptions wherein viral peaks are well represented by the model, the method may not yield an accurate prediction. As calibrated with only half the data in time, the model does not accurately predict the long-term off-treatment steady state exhibited by patient 26. The model fit resulting from using half the data underpredicts the viral load when off treatment, which is likely related to either an underestimation of viral productivity or viral infectivity, overestimation of viral death rate, or poor modeling/estimation of immune responsiveness. In this period of treatment discontinuation (beginning about 850 days into observation), the T-cells, virus, and immune responders all interact without the intervention of drugs, i.e., naturally. The prediction may therefore also be due to the overly simplistic and limited immune response model considered here. A conservative estimate of the off treatment
Figure 19: Model fit to data (‘x’) for patient 14 with parameters estimated from half longitudinal data (solid line) or full dataset (dash-dot line). Circles denote predictions of censored data measurements and the vertical line delineates between the two halves of the longitudinal data.
Figure 20: Model fit to data (‘x’) for patient 4 with parameters estimated from half longitudinal data (solid line) or full dataset (dash-dot line). Circles denote predictions of censored data measurements and the vertical line delineates between the two halves of the longitudinal data.
Figure 21: Model dynamics using parameters estimated from half of the time series data from patient 4.
viral load setpoint results even when using the full time series data, suggesting that for this patient, fixing the other parameters at the prescribed average values made it difficult to fit the model. This method should probably be accompanied by a measure of certainty of the prediction, perhaps based on how well the early time series data has been fit. (Although it would be difficult to argue which of the two fits to the early longitudinal data is “better”.)

As shown in Figure 23 for patient 24, it is possible to gain valuable information about a patient from even a single treatment interruption. Patients 10, 12, 13, and 25 yield similar results. Overall in most cases, having one or two treatment interruptions yields a good prediction of long-term viral dynamics. For example, for the eleven patients with a single treatment interruption during the first half of their data, the method only severely mispredicts the remaining data for patient 27. For the remainder of patients we see similar results when using half or all of the data.

As one might expect, if a patient does not undergo a therapy interruption during the observation period used to fit the model, it is difficult to predict a later treatment interruption. We examine this for patient 3 in Figure 24 and similar results hold for patients 23 and 47. The difference in model dynamics is reflected in the estimated parameters. For example, for patient 3, the estimates for $N_T$ (average virus released per burst T-cell) are $1.829e+01$ (half data) versus $3.677e+01$ (full data), so the underprediction of viral
Figure 23: Model fit to data (‘x’) for patient 24 with parameters estimated from half longitudinal data (solid line) or full dataset (dash-dot line). Circles denote predictions of censored data measurements and the vertical line delineates between the two halves of the longitudinal data.
Figure 24: Model fit to data ('x') for patient 3 with parameters estimated from half longitudinal data (solid line) or full dataset (dash-dot line). Circles denote predictions of censored data measurements and the vertical line delineates between the two halves of the longitudinal data.
load in this case may well be simply the result of underestimation of the rate $N_T$ when only using half the data.
20 Conclusions

A number of goals have been achieved in this paper. First, we have demonstrated that we can fit a complex mathematical model of HIV infection to long-term time series clinical data for individual patients. The data includes patients who experienced treatment interruptions. The novel inverse problem method we employ incorporates a censored data algorithm. After fitting the model to data, we investigated capabilities of the model in prediction. In particular our findings suggest that: (i) one may fit all time series and subsequently predict possible multiple stable steady states and (ii) one could use a subset of data to fit the model and then extrapolate over longer time horizons to predict viral load set points that might be most valuable in therapy decisions. We thus demonstrate that the treatment interruption data can provide crucial information for model fitting in terms of determining CD4 and viral load steady states.

While the model provides reasonable fits to most patient data, there are rather obvious areas for model improvement including additional compartments to better represent overall immune response to infection. We are currently pursuing such modeling efforts among others.
21 An Extended Model

A system of ordinary differential equations is formulated to describe the pathogenesis of HIV infection, wherein certain features that have been shown to be important by recent experimental research are incorporated in the model. These include the role of CD4+ memory cells that serve as a major reservoir of latently infected cells, a critical role for T-helper cells in the generation of CD8 memory cells capable of efficient recall response, and stimulation by antigens other than HIV. A stability analysis illustrates the capability of this model in admitting multiple locally asymptotically stable (locally a.s.) off-treatment equilibria. We show that this more biologically-detailed model can exhibit the phenomenon of transient viremia experienced by some patients on therapy with viral load levels suppressed below the detection limit. We also show that the loss of CD4+ T-cell help in the generation of CD8+ memory cells leads to larger peak values for the viral load during transient viremia. Censored clinical data is used to obtain parameter estimates. We demonstrate that using a reduced set of 16 free parameters, obtained by fixing some parameters at their population averages, the model provides reasonable fits to the patient data and, moreover, that it exhibits good predictive capability. We further show that parameter values obtained for most clinical patients do not admit multiple locally a.s off-treatment equilibria. This suggests that treatment to move from a high viral load equilibrium state to an equilibrium state with a lower (or zero) viral load is not possible for these patients.

22 Introduction

Since the seminal work of Ho, et al., [152] demonstrated the promise for elucidating HIV disease mechanisms through mathematical modeling, a wide variety of models have been proposed to describe various aspects of in-host HIV infection dynamics (e.g., [133, 134, 135, 144, 149, 159, 160]). The most basic of these models typically include two or three of the key dynamic compartments: virus, uninfected target cells, and infected cells. These compartmental depictions lead to systems of linear or nonlinear ordinary differential equations in terms of state variables representing the concentrations in each compartment and parameters describing viral production and clearance, cell infection and death rate, treatment efficacy, etc. Solutions for the model states yield the time course of viral load and CD4+ counts, for example. Although such models can be expected only to approximate the myriad processes underlying HIV pathogenesis, when used in conjunction with data as part of designed experiments, these models can be powerful tools in answering questions about the pathogenesis of HIV infection or similar biological processes. Mathematical models can also stimulate further clinical and laboratory research [160]. For example, early applications of linear systems to short-term data on patients undergoing ARV therapy suggested the now widely-held theory of very rapid and constant turnover of viral and infected cell
populations [152, 159], contradicting previous assumptions that stable viral and CD4+ concentrations during the clinical latency period of chronic HIV infection are due to absence of significant viral replication.

While the model developed and analyzed here is new, it modifies and extends both conceptually and structurally the predictive model in [135]. That model included both CD4+/viral dynamics as in models discussed in [144] as well as immune response compartments whose importance have been earlier established [142, 158, 173] – see the discussions in [133]. However, since development of that model, important features of HIV pathogenesis have emerged. In particular, the key role of CD4+ memory cells as a latent reservoir for HIV has been clearly established in the experimental literature [150, 167]. As the authors in [167] note, even in treated patients who have had no detectable viremia for as long as 7 years, the latent reservoir decays so slowly that early initiation of Highly Active Anti-Retroviral Therapy (HAART) with the goal of virus eradication is not likely to succeed. Another important feature of HIV pathogenesis that has emerged is the critical role for T-helper cells in the generation of CD8 memory cells capable of an efficient recall response [136, 140, 153]. In any discussions of mathematical modeling of complex systems it is appropriate to point out that while complex models may be needed to provide accurate descriptions of the underlying dynamics, the models are most useful when they can be compared to clinical and/or experimental data and can also be used for prediction. In developing models for HIV infection and treatment or some other biological phenomenon, this requires a balance between complexity and utility.

Hence, in this paper we do not try to formulate a model that reflects all features of cellular immune response as well as all host and viral factors. Instead, we attempt to develop a model that can capture the most salient biological features of disease progression, one for which parameters can be plausibly estimated based on clinical data over long observation periods with treatment interruptions, one that has predictive capabilities, and one for which control/drug therapy design is tractable (e.g., one that exhibits multiple equilibria). Clinical data includes the usual measurements of CD4+ T-cell count and censored viral load, as well as new data that we will collect, such as HIV-specific CD8+ T-cell and phenotypic data (discrimination between naive, memory and effector T-cells). The paper is organized as follows. In Section 23, a system of ordinary differential equations is developed to describe the pathogenesis of HIV and the cellular immune response. In Section ?? we discuss the ability of the model to admit multiple locally a.s. off-treatment equilibria. In Section ?? we demonstrate that one of the features of this more detailed biological model is the ability to exhibit the phenomenon of transient viremia following a non-HIV secondary infection, as proposed and demonstrated by others [154]. Furthermore, the phenomenon of transient viremia is used to elucidate the role of latently-infected CD4+ memory cells and the effect of CD4+ help on CD8+ memory during the ensuing immune response. In Section 24 the expectation maximization algorithm leading to weighted least-squares techniques is employed to fit the model.
to clinical data with lower limit censoring. The predictive capability of the model is also investigated by using simulation results, with parameters estimated from only half of the longitudinal observations, to predict the immune response in the latter half and comparing these predictions to clinical observations. In Section ?? we use the parameter estimates obtained in Section 24 to predict the number of off-treatment infected and un-infected equilibrium states available to each patient. Finally we close with conclusions and remarks in Section 25.

23 HIV Model

The model we develop in this paper conceptually modifies and extends the model in [135], wherein two types of target cells (CD4+ T-cells and macrophages), along with their corresponding infected states, free virus, and immune effector cells (CTL) are included in the model. Clinical data fitting results show that the preliminary model of [135] provides reasonable fits to most patient data and has impressive predictive capability when comparing model simulations, with parameters based on estimation using only half of the longitudinal observations, to the full longitudinal data sets. However, that model does not incorporate some important features of HIV pathogenesis and the cellular immune response, such as CD4+ memory cells as the major reservoir of latently infected cells and a critical role for T-helper cells in the generation of CD8 memory cells capable of an efficient recall response. To incorporate these important features, we thus seek a model that includes some measure of CD4+ T-helper cells, infected memory CD4+ T-cells and HIV-specific memory CD8+ T-cells. To retain the simplicity of the model, secondary target cells, such as macrophages, are not included as a compartment in our new model. It is worth noting that omitting the secondary target cells should not affect our clinical data fitting and predictive capabilities since this type of cell, even though it is very important at the beginning of infection, does not contribute significantly to the virus pool in the long run [151]. The model compartments are illustrated in Table 3, wherein the resting CD4+ T-cells \( T_2 \) are assumed to include naive CD4+ T-cells and memory CD4+ T-cells. This is reasonable since these two types of cells have similar behavior such as longer life spans and distribution in the lymphoid tissue. Once these resting CD4+ T-cells become activated, either through antigen priming of naive cells or reactivation of memory cells, they are more susceptible to HIV infection than resting cells and suffer elevated mortality [146]. Hence, we include these activated naive cells and reactivated memory cells in the other compartment as activated CD4+ T-cells \( T_1 \). Infected resting and activated cells are represented by the \( T_2^* \) and \( T_1^* \) states, respectively. A schematic of this new model is depicted in Fig. 25.

The corresponding compartmental ordinary differential equation (ODE) model for in-
<table>
<thead>
<tr>
<th>states</th>
<th>unit</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_1$</td>
<td>cells/\mu l-blood</td>
<td>uninfected activated CD4+ T-cells</td>
</tr>
<tr>
<td>$T'_1$</td>
<td>cells/\mu l-blood</td>
<td>infected activated CD4+ T-cells</td>
</tr>
<tr>
<td>$T_2$</td>
<td>cells/\mu l-blood</td>
<td>uninfected resting CD4+ T-cells</td>
</tr>
<tr>
<td>$T'_2$</td>
<td>cells/\mu l-blood</td>
<td>infected resting CD4+ T-cells</td>
</tr>
<tr>
<td>$V_I$</td>
<td>RNA copies/ml-plasma</td>
<td>infectious free virus</td>
</tr>
<tr>
<td>$V_{NI}$</td>
<td>RNA copies/ml-plasma</td>
<td>non-infectious free virus</td>
</tr>
<tr>
<td>$E_1$</td>
<td>cells/\mu l-blood</td>
<td>HIV-specific effector CD8+ T-cells</td>
</tr>
<tr>
<td>$E_2$</td>
<td>cells/\mu l-blood</td>
<td>HIV-specific memory CD8+ T-cells</td>
</tr>
</tbody>
</table>

Table 3: Model States.

Figure 25: Flow chart of model (51) with compartments as described in Table 25. Solid black arrows indicate death/clearance, solid gray arrows indicate birth/input. PI and RTI denote protease inhibitor and reverse transcriptase inhibitor, respectively.
host HIV infection dynamics is based on balance laws and is given by

\[
\begin{align*}
\dot{T}_1 &= -d_1 T_1 - (1 - \xi_1(t))k_1 V_t T_1 - \gamma T_1 + p_T \left( \frac{a_T V_t}{V_t + K_V} + a_A \right) T_2, \\
\dot{T}_1^* &= (1 - \xi_1(t))k_1 V_t T_1 - d_1 T_1^* - m E_1 T_1^* - \gamma T_1^* + p_T \left( \frac{a_T V_t}{V_t + K_V} + a_A \right) T_2^*, \\
\dot{T}_2 &= \lambda_T \frac{K_s}{V_t + K_s} + \gamma T_1 - d_2 T_2 - (1 - f \xi_1(t)) k_2 V_t T_2 - \left( \frac{a_T V_t}{V_t + K_V} + a_A \right) T_2, \\
\dot{T}_2^* &= \gamma T_1^* + (1 - f \xi_1(t)) k_2 V_t T_2 - d_2 T_2^* - \left( \frac{a_T V_t}{V_t + K_V} + a_A \right) T_2^*, \\
\dot{V}_I &= (1 - \xi_2(t))10^3N_T \sigma T_1^* - c V_I - 10^3[(1 - \xi_1(t)) \rho_1 k_1 T_1 + (1 - f \xi_1(t)) \rho_2 k_2 T_2] V_I, \\
\dot{V}_{NI} &= \xi_2(t)10^3 N_T \sigma T_1^* - c V_{NI}, \\
\dot{E}_1 &= \lambda E + \frac{b_{EI} T_1^*}{T_1^* + K_{EI}} E_1 - \frac{d_{EI} T_1^*}{T_1^* + K_{dEI}} E_1 - \delta E_1 E_1 - \gamma E \frac{T_1 + T_1^*}{T_1 + T_1^* + K_{\gamma}} E_1 + \frac{b_{EIE} V_t}{V_t + K_V} E_2, \\
\dot{E}_2 &= \gamma E \frac{T_1 + T_1^*}{T_1 + T_1^* + K_{\gamma}} E_1 + \frac{b_{E2} K_{EI}}{E_2 + K_{EI}} E_2 - \delta E_2 E_2 - \frac{a_T V_t}{V_t + K_V} E_2,
\end{align*}
\]

with an initial condition vector

\[
[T_1(0), T_1^*(0), T_2(0), T_2^*(0), V_I(0), V_{NI}(0), E_1(0), E_2(0)]^T.
\]

Here the factors $10^3$ are introduced to convert between microliter and milliliter scales, preserving the units from some of the earlier published papers [133, 142]. The treatment factors $\xi_1(t) = \epsilon_1 u(t)$ and $\xi_2(t) = \epsilon_2 u(t)$ represent the effective treatment impact, consisting of efficacy factors $\epsilon$ modeling the relative effectiveness of reverse transcriptase inhibitor (RTI), $\epsilon_2$ describing the relative effectiveness of protease inhibitor (PI), and a time-dependent treatment function $u(t)$ ($0 \leq u(t) \leq 1$) representing HAART drug level, where $u(t) = 0$ is fully off and $u(t) = 1$ is fully on. Since HIV treatment is nearly always administered as combination therapy, we do not consider the possibility of monotherapy, even for a limited period of time, though this could be implemented by considering separate treatment functions.

Since the $T_2$ compartment includes naive CD4+ T-cells, the input term $\lambda_T \frac{K_s}{V_t + K_s}$ for $T_2$ compartment is used to account for the source rate of naive CD4+ T-cells. This term depends on the viral load level since the thymus production can be diminished if the viral load is too high [156]. To limit the introduction of additional parameters, we assume that uninfected and infected resting CD4+ T-cells ($T_2$ and $T_2^*$, respectively) have the same natural death rate $d_2$. We remark that activated CD4+ T-cells have a higher natural death rate than resting memory and naive cells, and we use $d_1$ to denote the natural death rate of uninfected activated CD4+ T-cells $T_1$. The immune effector cells $E_1$ remove infected activated cells CD4+ T-cells $T_1^*$ from the system by cell lysis with a rate $m$. However, immune effector cells do not remove infected resting cells $T_2^*$, since these cells are in a quiescent state where the virus is not replicating and, thereby, escape
the detection of the immune effector cells. These infected resting cells are assumed to become targets for lysis only after activation [141].

The infected activated cells \( T^*_1 \) result from encounters between uninfected activated cells \( T_1 \) and free infectious virus \( V_I \) with infection rate \( k_1 \). The resulting term \( k_1 V_I T_1 \) is modified by a factor \( 1 - \xi_1(t) \) to account for RTI treatment. Infection of the resting T-cell compartment \( T_2 \), which is comprised of both memory and naive CD4+ T-cells, can occur in a number of ways. First, the most commonly transmitted R5 virus form of HIV-1 that utilize the chemokine receptor CCR5 can enter a subset of resting memory cells that express sufficient levels of CCR5 to support infection [141]. In addition, the X4 form of the virus can infect CD4+ T-cells, whether they belong to the naive or memory subsets. However, infection of naive and memory cells through these routes occurs much less frequently than infection of \( T_1 \), and, once infected, these cells often do not progress to a long-term stably-infected state in which the virus is integrated into the host DNA. In addition, it has been shown that infected naive CD4+ T-cells do not significantly contribute to the pool of infected resting CD4+ T-cells [143]. Hence, the term \( (1 - f\xi_1(t)) k_2 V_I T_2 \) is used to represent the infection process that results from encounters between the uninfected resting CD4+ T-cells and free virus \( V_I \), but with an infection rate \( k_2 < k_1 \) to account for a significantly lower rate of infection as compared to activated CD4+ T-cells. The treatment factor \( \xi_1(t) \) is potentially more effective in \( T_1 \) than in \( T_2 \), where the efficacy is modelled by \( f\xi_1(t) \) with \( 0 \leq f \leq 1 \).

A much more stable form of latent infection arises when activated CD4+ T-cells that have integrated HIV-1 DNA survive long enough to revert back to resting memory state, and latently infected resting CD4+ T-cells with integrated HIV-1 DNA are present in all infected individuals but only at low frequency [141, 147]. Hence, the terms involving \( \gamma_T T^*_1 \) are included in the model to account for the phenomenon of differentiation of infected activated CD4+ T-cells into infected memory or resting CD4+ T-cells \( T^*_2 \) at rate \( \gamma_T \). For simplicity, the rate at which uninfected activated CD4+ T-cells \( T_1 \) differentiate into uninfected resting CD4+ T-cells \( T_2 \) is also assumed to be \( \gamma_T \); the model could be extended easily to the case with different differentiation rates.

As the authors of [141] concluded, there is turnover (activation) in the latent reservoir when patients are viremic and the degree of turnover depends on the level of viremia. We thus assume that the activation of infected HIV-specific resting CD4+ T-cells depends on the virus concentration with a half-saturation constant \( K_V \). Hence, the terms involving \( \frac{a_T V_I}{V_I + K_V} T^*_2 \) are used to represent the activation of infected HIV-specific resting CD4+ T-cells with maximum activation rate of \( a_T \). Again to preserve the simplicity of this model, we assume that activation of uninfected HIV-specific resting CD4+ T-cells \( T_2 \) also depends on the virus concentration, with a half saturation constant \( K_V \), and that the maximum activation rate is also \( a_T \). Thus, the terms involving \( \frac{a_T V_I}{V_I + K_V} T_2 \) represent the activation of uninfected HIV-specific resting CD4+ T-cells. In order to incorporate the activation of resting CD4+ T-cells by some non-HIV antigen and preserve the simplicity of the model, we include the simple terms \( a_A T_2 \) and \( a_A T^*_2 \) into our model to describe this
phenomenon, with $a_A$ being the activation rate by non-HIV antigen. The parameter $a_A$ here can be utilized as a constant to represent a chronic level of infection or as a function of time $t$ to describe infections that are cleared by the body. These activation terms represent losses to the $T_2$ and $T^{\ast}_2$ compartments, with corresponding gain terms for the $T_i$ and $T^{\ast}_i$ compartments. However, the gain terms for $T_i$ and $T^{\ast}_i$ include a multiplicative factor $p_T$ to account for the net proliferation due to clonal expansion and programmed contraction. For simplicity, we assume that uninfected and infected CD4+ T-cells have the same expansion factor $p_T$; again this can be readily extended to include processes with different expansion factors.

Virus in the reservoir $T^{\ast}_2$ of infected resting CD4+ T-cells is latent and no virus can be produced by these cells unless they are activated [141]. Hence, free virus particles $V_I$ are produced only by activated infected CD4+ T-cells during viral budding leading up to viral produced lysis $\delta T^{\ast}_2$ of the CD4+ T-cells. The parameter $N_T$ accounts for the number of RNA copies produced during this process in the viral source term $\left(1 - \xi_2(t)\right)10^3 N_T \delta T^{\ast}_2$. In addition to a natural clearance rate $c$, we also include term $10^3[(1 - \xi_1(t))\rho_1 k_1 T_1 + (1 - f \xi_1(t))\rho_2 k_2 T_2]V_I$ in the free virus compartment $V_I$ to account for the removal of free virus that takes place when free virus infects $T_1$ and $T_2$. We make the simplifying assumption that $\rho_i = 1$ copies ml-blood cells ml-plasma, i.e., one free virus particle is responsible for each new infection. This could be adapted easily for multiple virus particles being responsible for each new infection by choosing $\rho_i > 1$. Since clinical measurements of viral load do not differentiate between infectious and non-infectious virus, we include a compartment in the model for tracking the amount of non-infectious virus $V_{NI}$ (viral load $V_L = V_I + V_{NI}$). The action of a protease inhibitor, resulting in the production of non-infectious virus $V_{NI}$ by infected cells is modeled by $\xi_2$. It should be noted that the inclusion of this additional state does not affect the dynamics of the other state variables, but its inclusion is necessary in order to use the clinical data for the total viral load $V_L$ in our inverse problem methodology below.

The source term $\lambda_{E_1}$, the constant death term $\delta_{E_1}$, and the nonlinear infected cell-dependent birth and death terms $\frac{b_{E_1} T_1^2}{T_1 + K_{b1}} E_1$ and $\frac{d_{E_1} T_1^2}{T_1 + K_{d1}} E_1$ terms in the $E_1$ compartment are adopted from the model in [134, 135], where the authors suggested that, by including such terms in the immune effector compartment, the model can admit multiple stable off-treatment steady states and exhibit transfer between “healthy” and “unhealthy” locally stable steady states via optimal or suboptimal structure treatment interruptions (STI) therapies. This makes it a good candidate for our investigation. Memory CD8+ T-cells are also subject to strict homeostatic control [168]; background expansion of memory cells through intermittent cell division being countered by an equivalent level of cell death. Hence, we include the term $\frac{b_{E_2} K_{b2}}{E_2 + K_{b2}} E_2 - \delta_{E_2} E_2$ for homeostatic regulation in the $E_2$ compartment, similar to that used in [171]. In the homeostatic regulation, $b_{E_2}$ represents the maximum proliferation rate and $\delta_{E_2}$ corresponds to the death rate, where the proliferation signal decreases linearly with population size.
The term $\gamma_E \frac{T_1 + T_*}{T_1 + T_* + K_1} E_1$ in the model is used to include the essential role that activated CD4+ T-cells play in the generation of memory CD8+ T-cells, where parameter $K_1$ is a half-saturation constant and $\gamma_E$ is the maximum rate at which $E_1$ differentiates into $E_2$. Since depletion of CD4+ cells has a minimal effect during the recall response [166, 172], the term $\frac{aE_1V}{r_k + K_V} E_2$ for reactivation of memory CD8+ T-cells is independent of CD4+ T-cell help. Similar to the activation of HIV-specific resting CD4+ T-cells, we assume that activation of HIV-specific memory CD8+ T-cells also depends on the virus concentration. For simplicity, we use the same half-saturation constant $K_V$ for the activation of memory CD8+ T-cells. Since CD8+ T-cells tend to divide sooner and to have a faster rate of cell division than CD4+ T-cells [165], we use a different parameter $p_E$ to account for the net proliferation due to clonal expansion and programmed contraction of activated CD8+ T-cells in the $E_1$ compartment.
24 Application of the Model to Clinical Data

The data for our investigations come from Massachusetts General Hospital (MGH), where all the patients enrolled in the study are symptomatic with acute or early HIV-infection (for more detailed information of these data, the interested reader is referred to [135, 155, 161]). In summary, nearly all subjects in the study undergo combination therapy and many have at least one treatment interruption. The available clinical data include total CD4+ T-cell count and total RNA copies, where for model (51) the total CD4+ T-cell counts are represented by \( \bar{z}_1(t; q) = T_1(t; q) + T_1^*(t; q) + T_2(t; q) + T_2^*(t; q) \) and total RNA copies are represented by \( \bar{z}_2(t; q) = V_I(t; q) + V_{NI}(t; q) \). If the measurements of RNA copies are below the limit of quantification for the assay used (400 copies/ml-plasma for a standard assay and 50 copies/ml-plasma for an ultra-sensitive assay), then the observed viral load value is censored to be at its detection limit; that is, in these cases the observed values do not represent the true data values anymore. Furthermore, observations of viral load and CD4+ may not be at the same time points and the observation times and intervals vary substantially among patients. So, in general, for patient number \( j \) we have CD4+ T-cell data pairs \((t_{ij}^1, \bar{y}_{ij}^1)\), \( i = 1, \ldots, N_j^1 \), and potentially different time point viral RNA data pairs \((t_{ij}^2, \bar{y}_{ij}^2)\), \( i = 1, \ldots, N_j^2 \). Hence, the clinical data for carrying out the inverse problem involves partial observations, measurements from combined compartments, and highly censored viral load measurements.

We use the clinical data in two ways: for estimating parameter values and for testing the predictive capability of the model. For the (new) model (51) presented here most parameters can not be determined \textit{a priori} and therefore must be estimated from data. In Section 24.1 we show how we use the clinical data to formulate an inverse problem to obtain parameter estimates for each patient. Using the collective parameter estimates for all the patients and knowledge of model sensitivities we then fix some model parameters and initial conditions at the \textit{population averages} across these patients and re-estimate the remaining free parameters for each individual.

In Section 24.2 we use the same values for the fixed model parameters and initial conditions as obtained in Section 24.1 and estimate the remaining free parameters using only \textit{half} of the longitudinal data for each patient. Using the parameters thus obtained, we use the model to predict the time course of the disease during the time period corresponding to the second half of the longitudinal data and compare these model predictions to the actual data.

As one might expect, if a patient does not have a sufficient number of observations or does not undergo a therapy interruption during the observation period used to fit the model, then it is difficult to obtain accurate estimations of the dynamically dependent parameters. We are interested in using this model to predict later disease progression involving both therapy and treatment interruption (see Section 24.2). Hence, for our purposes we focus on 14 patients with sufficient data and at least one on/off treatment schedule in the first half of their longitudinal data. Note that for each of these patients,
24.1 Fitting the Model/Parameter Estimation

To obtain patient specific-parameter estimates for the model, we use individual patient data and carry out an inverse problem. The technique we use in this paper is adapted from the one in [135], where the authors developed a statistically-based censored data method (an expectation maximization algorithm) combined with an ordinary nonlinear least-squares technique. We note that the variance \( \sigma_r^2 \) in CD4 measurements and the variance \( \sigma_v^2 \) in viral load measurements are likely to be different due to assay differences. Hence, in this paper we use the expectation maximization algorithm based on Maximum Likelihood Estimation for \((q, \sigma_r^2, \sigma_v^2)\) which, under normality assumptions on the errors, results in a weighted least-squares technique (see [164]) with solution given by

\[
\hat{q}^j = \arg \min_{q \in \mathbb{Q}} \left[ \frac{1}{\hat{\sigma}_r^2} \sum_{i=1}^{N^j_1} |y_{ij}^1 - z_1(t_{ij}^1; q)|^2 + \frac{1}{\hat{\sigma}_v^2} \sum_{i=1}^{N^j_2} |y_{ij}^2 - z_2(t_{ij}^2; q)|^2 \right]
\]

\[
\hat{\sigma}_k^2 = \frac{1}{N^j_k} \sum_{i=1}^{N^j_k} |y_{ik}^j - z_k(t_{ik}^j; \hat{q}^j)|^2, \quad k = 1, 2,
\]

for the log\(_{10}\)-transformed system of model (51) for patient \( j \), where \( y_{ij}^1 = \log_{10} \bar{y}_{ij}^1 \) and \( z_1(t_{ij}^1; q) = \log_{10}(\bar{z}_1(t_{ij}^1; q)) \), \( i = 1, \cdots, N^j_1 \), and \( y_{ij}^2 = \log_{10} \bar{y}_{ij}^2 \), \( z_2(t_{ij}^2; q) = \log_{10}(\bar{z}_2(t_{ij}^2; q)) \), \( i = 1, \cdots, N^j_2 \). As noted in [135], by using a log-transformed system one can resolve a problem of states becoming unrealistically negative due to round-off error: nonnegative solutions of this model should stay so throughout numerical simulation. This approach also enables efficient handling of unrealistic cases where states get infinitesimally small during integration due to parameters selected by optimization algorithms. From a statistical point of view, log transformation is a standard technique to render the observations more nearly normally distributed, which also supports use of the weighted least squares criterion as an equivalent to maximum likelihood estimation.

The expectation maximization (EM) algorithm is outlined below. To simplify the notation, we drop the patient index \( j \) in this algorithm description. The following notation will be used in the algorithm: the relevant censoring point at time \( t^i \) is represented by \( L^i \) and \( \phi^i \) is the indicator function for the set \( \{y_2^i > L^i\} \), \( \phi \) denotes the standard normal probability density function and \( \Phi \) is the corresponding cumulative distribution function. For each patient, we carried out the following parameter estimation algorithm:

- (Step 1) Create adjusted data \( \tilde{y}^i \) by replacing censored \( y_2^i \) values by \( L^i / 2 \), and use ordinary least squares to estimate \( \tilde{q}^{(0)} \) using both CD4+ data \( y_1^i \) and adjusted viral

\[ N^j_1 \] is not substantially different than \( N^j_2 \).
RNA data $\tilde{y}^i$ (which now includes replaced censored values).

$$q^{(0)} = \arg \min_{q \in Q} \left[ \sum_{i=1}^{N_1} |y^i_1 - z_1(t^i_1; q)|^2 + \sum_{i=1}^{N_2} |\tilde{y}^i - z_2(t^i_2; q)|^2 \right].$$

Obtain an initial estimate for $\sigma_1^2$ and $\sigma_2^2$ from

$$\left(\hat{\sigma}_1^{(0)}\right)^2 = \frac{1}{N_1} \sum_{i=1}^{N_1} |y^i_1 - z_1(t^i_1; \hat{q}^{(0)})|^2, \quad \left(\hat{\sigma}_2^{(0)}\right)^2 = \frac{1}{N_2} \sum_{i=1}^{N_2} |\tilde{y}^i - z_2(t^i_2; \hat{q}^{(0)})|^2,$$

respectively. Set $k = 0$.

- (Step 2) Define $\hat{z}_2^{(k)} = z_2(t^i_2; \hat{q}^{(k)})$ and $\hat{\xi}^{(k)} = \frac{L^i - \hat{z}_2^{(k)}}{\hat{\sigma}_2^2}$. Update the data and residuals by

$$\hat{y}^{(k)} = \chi^i y^i_2 + (1 - \chi^i) \left[ \hat{z}_2^{(k)} - \hat{\sigma}_2^{(k)} \frac{\phi(\hat{\xi}^{(k)})}{\phi(\hat{\xi}^{(k)})} \right],$$

$$\hat{r}^{(k)} = \chi^i (y^i_2 - \hat{z}_2^{(k)})^2 + (1 - \chi^i) (\hat{\sigma}_2^{(k)})^2 \left[ 1 - \frac{\hat{\xi}^{(k)} \phi(\hat{\xi}^{(k)})}{\phi(\hat{\xi}^{(k)})} \right].$$

- (Step 3) Update the estimate of $q$ to $\hat{q}^{(k+1)}$ by performing the weighted least squares minimization in the parameters $q$

$$\hat{q}^{(k+1)} = \arg \min_{q \in Q} \left[ \frac{1}{(\hat{\sigma}_1^{(k)})^2} \sum_{i=1}^{N_1} |y^i_1 - z_1(t^i_1; q)|^2 + \frac{1}{(\hat{\sigma}_2^{(k)})^2} \sum_{i=1}^{N_2} |\hat{y}^{(k)} - z_2(t^i_2; q)|^2 \right],$$

and computing

$$\left(\hat{\sigma}_1^{(k+1)}\right)^2 = \frac{1}{N_1} \sum_{i=1}^{N_1} |y^i_1 - z_1(t^i_1; \hat{q}^{(k+1)})|^2, \quad \left(\hat{\sigma}_2^{(k+1)}\right)^2 = \frac{1}{N_2} \sum_{i=1}^{N_2} \hat{r}^{(k)}.$$

If relative changes in $\hat{q}$, $\hat{\sigma}_1$ and $\hat{\sigma}_2$ are small, terminate. Otherwise set $k = k + 1$ and go to Step 2.

Although the EM algorithm may seem unnecessarily complicated, it should be noted that previous attempts to fit censored clinical data without the EM algorithm produced extremely poor results. For more details about the EM algorithm and how to carry it out, interested readers are referred to [135, 145, 157].

As noted previously, model (51) modifies and extends the model in [135] to incorporate important features of HIV pathogenesis. However, the cost of this more biologically detailed model is a larger set of free parameters: 31 model parameters and 8 initial conditions. Hence, we seek to reduce the number of free parameters. We first estimate
all 39 parameters for each of the 14 patients by applying the EM algorithm to the full longitudinal data set, where the total number of data points (sum of number of CD4+ data points and number of viral load data points) varies from 71 to 273. We then fix 23 parameters at the population averages across these patients, and re-estimate the remaining 16 parameters for each patient by applying the EM algorithm to the full longitudinal data set. Note that there exists biological variation in all parameters across the patients, and there also exist high correlations among some of these parameters such as the RNA copies produced per infected cells $N_T$ and the virus natural death rate $c$. We also observe that sensitivity with respect to some of these parameters may be highly time-dependent. For example, the dynamics of the model is much more sensitive to the treatment efficacies $\epsilon_1$ and $\epsilon_2$ in the treatment periods than it is in the off-treatment periods. All of these considerations make it difficult to choose a priori which parameters can be fixed. In this paper, we empirically chose some parameters (such as the saturation parameters) to which the model appears to be relatively insensitive to take as fixed. Table 4 specifies all the fixed parameters (19 model parameters and 4 initial conditions) and their corresponding values.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_T$</td>
<td>3.792e-04</td>
</tr>
<tr>
<td>$k_2$</td>
<td>2.005e-09</td>
</tr>
<tr>
<td>$c$</td>
<td>5.818e+00</td>
</tr>
<tr>
<td>$K_{b1}$</td>
<td>2.488e-02</td>
</tr>
<tr>
<td>$\delta_{E1}$</td>
<td>5.967e-02</td>
</tr>
<tr>
<td>$K_\gamma$</td>
<td>1.357e+00</td>
</tr>
<tr>
<td>$K_s$</td>
<td>2.789e+04</td>
</tr>
<tr>
<td>$E_1^0$</td>
<td>6.821e-02</td>
</tr>
<tr>
<td>$d_2$</td>
<td>3.096e-03</td>
</tr>
<tr>
<td>$\delta$</td>
<td>2.095e-01</td>
</tr>
<tr>
<td>$\lambda_E$</td>
<td>9.930e-04</td>
</tr>
<tr>
<td>$d_E$</td>
<td>6.278e-02</td>
</tr>
<tr>
<td>$K_{b2}$</td>
<td>8.697e+01</td>
</tr>
<tr>
<td>$K_v$</td>
<td>1.479e+01</td>
</tr>
<tr>
<td>$T_2^0$</td>
<td>7.521e-03</td>
</tr>
<tr>
<td>$V_{NI}^0$</td>
<td>3.571e+03</td>
</tr>
<tr>
<td>$f$</td>
<td>5.068e-01</td>
</tr>
<tr>
<td>$m$</td>
<td>1.127e-03</td>
</tr>
<tr>
<td>$b_{E1}$</td>
<td>3.885e-02</td>
</tr>
<tr>
<td>$K_d$</td>
<td>1.200e-01</td>
</tr>
<tr>
<td>$\gamma_E$</td>
<td>5.154e-04</td>
</tr>
<tr>
<td>$K_b$</td>
<td>2.488e-02</td>
</tr>
<tr>
<td>$d_E$</td>
<td>5.967e-02</td>
</tr>
<tr>
<td>$K_{b2}$</td>
<td>2.005e-09</td>
</tr>
<tr>
<td>$K_v$</td>
<td>6.278e-02</td>
</tr>
<tr>
<td>$T_2^0$</td>
<td>7.521e-03</td>
</tr>
<tr>
<td>$V_{NI}^0$</td>
<td>3.571e+03</td>
</tr>
</tbody>
</table>

Table 4: Average model parameter values (19) and initial conditions (4) used in model fitting with half and full longitudinal data set.

After we estimate the reduced set of 16 free parameters (using the full set of data, in an inverse problem), we simulate the CD4+ T-cell and viral load trajectories over the full time span of the patient’s observations by using the parameter values obtained and compare to the experimental data. Figures 26-29, and Figures 30-39 in the Appendix illustrate model fits (dash-dot lines) for all 14 patients. These figures reveal that the fits to CD4+ T-cells and the fits to viral load are reasonable when using the full data set to estimate the parameter values, as illustrated in Figure 26 for patient 1.
Figure 26: Clinical data (‘x’) for CD4 T-cells (upper graph) and viral load (lower graph). Also included are the model results with parameters estimated from full data (dash-dot line) or half longitudinal data (solid line). Dark and gray circles denote the predicted censored data values for the full and half longitudinal data, respectively. The vertical line in the middle of the graph delineates between the two halves of the longitudinal data. A solid line along the x-axis in the upper figure indicates periods when the patient is on HAART treatment, while a dashed line indicates off-treatment periods.

24.2 Predictive Capability of the Model

When proposing a new model, a necessary but often difficult requirement is validation of the model with clinical data. One way to do this is to use the first part of the available longitudinal data for a given patient to estimate patient specific parameters. One then uses these parameters in model simulations to accurately predict the subsequent progression of the disease as represented in the remainder of the longitudinal data for that patient. In this section, we retain the same values for the 23 fixed parameters as above (Table 4) and estimate the remaining 16 free parameters using only half of the longitudinal data, where the total number of data points varies from 42 to 154. Using the parameters thus obtained, we use model (51) to predict the time course of the CD4+ T-cell count and viral load corresponding to the second half of the longitudinal data and compare these model predictions to the actual data.

Figures 26-29, and Figures 30-39 in the Appendix illustrate model predictions (solid lines) for all 14 patients. These figures reveal that both the fits to CD4+ T-cells and the fits to viral load are reasonable when using half data to estimate the free parameter values, even with multiple on/off-treatments, as illustrated in Figure 27 for patient 4.
Figure 27: Clinical data (‘x’) for CD4 T-cells (upper graph) and viral load (lower graph). Also included are the model results with parameters estimated from full data (dash-dot line) or half longitudinal data (solid line). Dark and gray circles denote the predicted censored data values for the full and half longitudinal data, respectively. The vertical line in the middle of the graph delineates between the two halves of the longitudinal data. A solid line along the x-axis in the upper figure indicates periods when the patient is on HAART treatment, while a dashed line indicates off-treatment periods.

However, there is under- or over-prediction for some patients when using parameters estimated from the half data set. For example, model fits obtained by using half time series data for patient 2 are shown in Fig. 28 (solid line), where it can be seen that by using parameters estimated from the first half data set, the predicted values for viral load data are higher and the predicted values for CD4+ T-cell are lower than the observed data values, even though there are two off-treatment periods in the first half of the data set. This discrepancy may be due to the fact that both off-treatment periods are short and, therefore, provide very little information about the off-treatment behavior. Note that, for this patient, the values of most observations of viral loads in the off-treatment periods during the first half data set are higher than those observed in the second half data set and this makes it difficult to correctly predict future off-treatment trends. Hence, it is not just the number of off-treatment periods in the first half data set but also the length of each off-treatment period and the number of observations in these periods that determines the accuracy of the predictions of off-treatment viral load levels.
Figure 28: Clinical data (‘x’) for CD4 T-cells (upper graph) and viral load (lower graph). Also included are the model results with parameters estimated from full data (dash-dot line) or half longitudinal data (solid line). Dark and gray circles denote the predicted censored data values for the full and half longitudinal data, respectively. The vertical line in the middle of the graph delineates between the two halves of the longitudinal data. A solid line along the $x$-axis in the upper figure indicates periods when the patient is on HAART treatment, while a dashed line indicates off-treatment periods.

In contrast, Figure 29 compares the model fit and prediction obtained by using the half data set to clinical data for patient 10. Even though there is only one off-treatment period in the first half data set, the period is longer with a richer data set and the half data predictions of both CD4+ T-cell and viral load agree quite well in the second half data set. Hence, the results demonstrated in Figures 28 and 29 suggest that in order to have accurate predictive capability we need to obtain a sufficient number of representative data points in the time period providing the information about the dynamics (the time period used to estimate parameters).

Finally, we note that the model developed and tested here retains the predictive capabilities present in the earlier model of [135] in that the ability to predict subsequent data after fitting the model with half of the data is at least as good as that exhibited in [135] for the various patients in our data sets. In addition, the added complexity of creating a more accurate description of the underlying biological processes is greatly offset by fixing many of the parameter values at their population averages. While by no means a definitive validation of the model, these results suggest that, given data on a new patient, it may be possible to fit this new data with the smaller subset of 16 free
Figure 29: Clinical data (‘x’) for CD4 T-cells (upper graph) and viral load (lower graph). Also included are the model results with parameters estimated from full data (dash-dot line) or half longitudinal data (solid line). Dark and gray circles denote the predicted censored data values for the full and half longitudinal data, respectively. The vertical line in the middle of the graph delineates between the two halves of the longitudinal data. A solid line along the $x$-axis in the upper figure indicates periods when the patient is on HAART treatment, while a dashed line indicates off-treatment periods.

parameters (and fixed parameters in Table 4) and that, using the parameter values thus obtained, we can reasonably predict subsequent disease progression in the near future.
25 Conclusion and Remarks

We present a system of ordinary differential equations to describe the pathogenesis of HIV infection. This model provides a more accurate biological description of the underlying infection and immune response than a previous model [135], including a compartment for CD4+ memory cells that serve as a major reservoir of latently infected cells, a critical role for T-helper cells in the generation of CD8 memory cells capable of efficient recall response, and stimulation by antigens other than HIV. A stability analysis illustrates the capability of this model for admitting multiple locally asymptotically stable (locally a.s.) off-treatment equilibria. The existence of these off-treatment equilibria allow for the possibility of treatment strategies that would shift patients from higher viral load equilibrium states to lower viral load equilibrium states.

We show that this more biologically-detailed model can exhibit the phenomenon of “viral blips” or transient viremia experienced by some patients on therapy with viral load levels suppressed below the detection limit, something not possible with the previous model. We also use this more biologically-detailed model to explore the effect of CD4+ T-cell help on the CD8+ T-cell immune response in the context of transient viremia induced by a non-HIV infection. The results of this analysis demonstrates that in our model, impaired differentiation to CD8+ memory results in an infected equilibrium exhibiting a higher viral load and leads to a larger viral blip following a secondary infection.

Application of this model to censored clinical data to obtain parameter estimates demonstrates that using a reduced set of 16 free parameters, obtained by fixing 23 parameters at their population averages, the model provides good fits to the patient data. Furthermore, we demonstrate the strong predictive capability of the model for longer term behavior by using estimates of the 16 free parameters obtained from half of the longitudinal data to compare model predictions to the second half of the clinical data. These results suggest that it may be possible to fit new patient data with the smaller subset of 16 free parameters (and the fixed parameters in Table 4) and that, using the parameter values thus obtained, we may be able to predict subsequent disease progression in the near future. We further show that parameter values obtained for most clinical patients do not admit multiple locally a.s off-treatment equilibria in the corresponding model.

One of the challenges in our clinical data fitting is that we do not have sufficiently rich data sets to fit all the model parameters and initial conditions. Due to the high correlations between the model parameters, biological variations in all parameters across population and the sensitivity with respect to parameters varying over time (especially during the transition time between the off-treatment period and on-treatment period), we empirically chose some parameters in this paper to be fixed. Even though the simulation results indicate reasonable fits to data of each of the patients, we need to develop
a scientific methodology to deal with this situation in order to obtain more reliable parameter values to be fixed a priori and thereby sharpen our estimation results. As outlined above we reduced the number of parameters to be estimated by fixing some of the parameters at population averages. In clinical cases where large, long term individual patient longitudinal data sets are not available as in our study here, one could use population averages accrued in prior patient trials to reduce the number of free parameters. Another possible method to reduce the dimension of parameter space and the associated high correlation between these parameters is principal component analysis, which is currently under investigation.

The clinical data fitting results in Section 24 indicate that if one does not have a sufficient number of observations during the periods where the dynamics are changing, then it is difficult to obtain parameter estimates useful in predicting corresponding future trends in disease progression. Hence, the goodness of fit results are affected by not only the number of observations but also the sampling times for data collection. This general principle was also observed and explained conceptually in [138] where the general “information content” in data and its relation to sensitivity is explored. These considerations suggest that parameter sensitivity with respect to data is important in experimental design, assisting in reducing effort and resources required to collect necessary data. One potential approach for this is the generalized sensitivity function methodology as proposed in [170] and explored in [137, 139]. This approach combines the sensitivities of model output with respect to model parameters with the sensitivities of parameters estimates with respect to changes in model outputs. Obtaining results using this methodology is challenging even for simple examples and our initial efforts have not yet proved fruitful for our complex HIV models which feature high correlations between some parameters and dynamics that dramatically change during the transition time between off-treatment and on-treatment. Hence, one of our future efforts involves development of a methodology to determine improved sampling times for data collection. We are optimistic that ideas contained in principal component analysis and generalized sensitivity functions can be combined to provide new guidance in this regard.
Appendix: Section 24 Results for Other Patients

The model fits for the remaining clinical patients are shown in Figures 30-39. As in previous figures, clinical data (‘x’) for CD4 T-cells and viral load are shown in the upper and lower graphs, respectively. Also included in the figures are the model results with parameters estimated from full data (dash-dot line) or half longitudinal data (solid line). Dark and gray circles denote the predicted censored data values for the full and half longitudinal data, respectively. The vertical line in the middle of the graph delineates between the two halves of the longitudinal data. A solid line along the $x$-axis in the upper figure indicates periods when the patient is on HAART treatment, while a dashed line indicates off-treatment periods.

Figure 30: CD4 T-cells (upper graph) and viral load (lower graph).
Figure 31: CD4 T-cells (upper graph) and viral load (lower graph).

Figure 32: CD4 T-cells (upper graph) and viral load (lower graph).
Figure 33: CD4 T-cells (upper graph) and viral load (lower graph).

Figure 34: CD4 T-cells (upper graph) and viral load (lower graph).
Figure 35: CD4 T-cells (upper graph) and viral load (lower graph).

Figure 36: CD4 T-cells (upper graph) and viral load (lower graph).
Figure 37: CD4 T-cells (upper graph) and viral load (lower graph).

Figure 38: CD4 T-cells (upper graph) and viral load (lower graph).
Figure 39: CD4 T-cells (upper graph) and viral load (lower graph).
26 Size-Structured Population Models

27 Introduction: A Motivating Application

The mosquitofish, *Gambusia affinis*, is used throughout the world to control mosquito populations. Indigenous to the southeastern United States and northeastern Mexico, it is now one of the most widely distributed of all freshwater fish. When introduced into a rice field, the mosquitofish eat the water-borne mosquito larvae. Consequently, it is thought to be the most widely disseminated natural predator as well as the most popular form of mosquito control.

In spite of their widespread use, the mechanisms underlying the growth of Gambusia populations (and consequently, mosquito control) are not well understood. For example, studies have shown that application of Gambusia early in the rice season leads to fewer mosquito larvae on the average over several fields. However, there is considerable variability among rice fields, with some unstocked fields having fewer larvae than stocked fields.

In the early 1980’s a research group from UC-Davis [63] carried out experiments to better understand how Gambusia populations develop in rice fields. Their goal was to achieve better mosquito management through more detailed knowledge of Gambusia population and predation dynamics. Even though the economic implications were substantial, no one really knew how many mosquitofish should be used to stock a rice paddy field. In addition, stocking methods do significantly differ, raising many questions. For example, should all the mosquitofish be added initially, or should they be introduced into the rice paddy field periodically or by some other time dependent schedule?

There are a number of avenues that can be taken to investigate these questions. A control theorist might try to use a general system of ordinary differential equations such as

$$\dot{x} = Ax + Bu$$

and choose a control $u$ (stocking rate perhaps) to improve system behavior (see Chapter 7 for an introduction to the control theory). However, this requires knowing the matrices $A$ and $B$. At one time control theorists thought biologists might be able to provide $A$ and $B$, but they unfortunately were not able to do this with any degree of certainty.

Another avenue is to perform many experiments in hope of finding some empirical relationship. The approach that we pursue here is to adapt some sort of reasonable mathematical model to understand the basic dynamics of growth and decline in the mosquitofish population. Several types of population models have been developed over the years to model population dynamics. These include single species models, logistic models, predator/prey models and structured models, each of which will be discussed in the following sections.
The simplest population models are the single species models. Let \( p(t) \) denote the population (number) of a given species at time \( t \). Assuming that this population is isolated (that is, there is no net immigration nor emigration), then the rate of change of the population is simply the difference between the birth rate and the death rate:

\[
\frac{dp}{dt} = \text{birth rate} - \text{death rate}.
\]

We further assume that the more individuals there are, the more births and deaths that occur. That is, both the birth rate and death rate are proportional to the number of individuals in the population. Consequently, the birth rate is given by \( \beta p \) and death or mortality rate is \( \mu p \). In this case, the model becomes

\[
\frac{dp}{dt} = \beta p - \mu p = \alpha p, \tag{53}
\]

where \( \alpha = \beta - \mu \) represents the net rate of birth/death per individual in the population. Equation (53) is a linear first order differential equation and is known as the Malthusian law of population growth. If the population of a given species is \( p_0 \) at time \( t = t_0 \), then the solution to the initial-value problem has the form \( p(t) = e^{\alpha(t-t_0)}p_0 \). Depending on the value of \( \alpha \) the solution \( p(t) \) will have one of the following three characteristics: (i) when \( \alpha > 0 \) (more births than deaths) the population will grow exponentially with time, (ii) when \( \alpha \) is negative the population will die out, and (iii) when \( \alpha \) is equal to zero the population will remain constant and is equal to the initial number of individuals \( p_0 \) (see Figure 40).

Figure 40: Graphs of the population \( p(t) \).

The single species model is so simple that it predicts population outcomes that are clearly unreasonable. Note that the deaths in this model are from “natural causes” or old age. There is no predatory or otherwise harmful activities represented in this model.
Moreover, when the number of individuals $p$ becomes very large, the single species model cannot be very accurate, since it does not reflect the fact that individual members are now competing with each other for limited living space, natural resources, and food.

## 29 The Logistic Model

Clearly overcrowding will reduce the amount of food, as well as tax other resources such as oxygen levels, etc. In the single species model we can add a crowding term, which will result in more deaths with higher numbers of individuals. A simple first assumption might be that the death rate per individual $\mu$, is a function of the population $p$. That is, we might take $\mu = \mu(p)$. The simplest form of such a function is linear $\mu(p) = \mu p$, so that the model becomes

$$\frac{dp}{dt} = \beta p - (\mu p)p.$$  \hspace{1cm} (54)

This equation was first introduced by the Dutch mathematical biologist Verhulst in 1837 and has subsequently become known as the logistic equation. The term $\mu p$ in equation (54) simply translates to more deaths occurring when $p$ is large; this is the competition or crowding term.

We observe that if $p$ is small, $-\mu p^2$ is negligible and the model reduces to the Malthusian law. On the other hand, if $p$ is large, $-\mu p^2$ serves to slow down the rapid rate of increase. In either case, for $\mu \neq 0$, the equation is readily solved analytically via standard techniques. Using the method of separation of variables, we rewrite the differential equation

$$\frac{dp}{\beta p - \mu p^2} = dt.$$  \hspace{1cm} p(t_0) = p_0$$

as

$$\frac{dp}{\beta p - \mu p^2} = dt.$$  \hspace{1cm} p(t_0) = p_0$$

Hence we find

$$p(t) = \frac{\beta p_0}{\mu p_0 + (\beta - \mu p_0)e^{-\beta(t-t_0)}},$$

the graph of which is depicted in Figure 41. This solution is often written as

$$p(t) = \frac{Kp_0}{p_0 + (K-p_0)e^{-\beta(t-t_0)}}.$$
corresponding to the equation being written as
\[
\frac{dp}{dt} = \beta p \left(1 - \frac{p}{K}\right) = rp \left(1 - \frac{p}{K}\right),
\]
where \(K = \beta/\mu\) is the population’s *carrying capacity* and \(\beta = r\) is called the *intrinsic growth rate*. In terms of *per capita growth* we have
\[
\frac{1}{p} \frac{dp}{dt} = r \left(1 - \frac{p}{K}\right),
\]

![Graph of the solution to the logistic model.](image)

We remark that regardless of the initial population \(p_0\), the number of individuals always approaches the limiting value \(K = \beta/\mu\) as \(t \to \infty\). Furthermore, since
\[
\frac{d^2p}{dt^2} = \frac{d}{dt} \left(\frac{dp}{dt}\right) = \frac{d}{dt} (\beta p - \mu p^2) = \beta \frac{dp}{dt} - 2\mu p \frac{dp}{dt} = (\beta - 2\mu p)(\beta - \mu)p,
\]
it follows that if \(p < \frac{\beta}{2\mu}\), then \(\frac{d^2p}{dt^2} > 0\) and \(p\) is thus concave up. On the other hand, if \(p > \frac{\beta}{2\mu}\) (and \(p < \frac{\beta}{\mu}\)), then \(\frac{d^2p}{dt^2} < 0\) and \(p\) is concave down. Hence, the graph of \(p\) has the
form as depicted in Figure 41. Such a curve is called a *logistic*, or *S-shaped* curve. From its shape, the time period before the population reaches $\frac{\beta}{2\mu}$ is known as the period of accelerated growth. After this period, the rate of growth decreases and asymptotically reaches zero. This is a period of diminishing growth.

The logistic model is sometimes also called the Verhulst-Pearl model (it was developed by Verhulst [72] and later rediscovered and popularized by Pearl [70]). It has been widely used [68] for many years in certain applications. Its primary feature, the population saturation, is biologically realistic if nothing else is preying on the population. However, this model is not adequate in a predator/prey situation.

30 A Predator/Prey Model

In the mid-1920s, Italian biologist Umberto d’Auona studied the percent of total catch of selachians (a group of fish comprising the sharks, skates, and rays) in the Mediterranean port of Port Fiume, Italy. The data is tabulated for the period from 1914 to 1923 [64].

He was puzzled by the very large increase of selachians during World War I (1914-1918). He reasoned that selachians increased due to the reduced level of fishing during the war. Therefore, there were more fish available as food for the selachians, and hence the selachian population multiplied. However, this explanation was not satisfactory since one did not have more food fish (supposedly to be eaten by sharks) during this period. After exhausting all biological explanations, in 1926 [73] he turned to his colleague, the famous Italian mathematician Vito Volterra, for help. Volterra formulated a mathematical model for the growth of selachians and their prey, food fish, by separating all food fish into the prey population and selachians into the predator population.

Let the number of predators and prey at time $t$ be $N(t)$ and $E(t)$ (the edibles), respectively. A simple assumption is that the population of edibles will grow exponentially without the predators. In addition, the prey death rate depends on both $E$ and $N$ (since they are eaten by predators). Similarly, since the predators need the edibles to live, their birth rate will be dependent on $E$ and $N$ as well. Finally, with no edibles, the predators are assumed to die out exponentially. Then, we can write the following system of differential equations for the predator/prey model:

$$\frac{dN}{dt} = (\beta_N E)N - \mu_N N,$$

$$\frac{dE}{dt} = \beta_E E - (\mu_E N)E. \quad (55)$$

The system of equations (55), which is also called the Lotka-Volterra model, has two equilibrium solutions:

$$N^e = E^e = 0$$
and 

\[ N^e = \frac{\beta_N}{\mu_N}, \quad E^e = \frac{\mu_N}{\beta_N} \]

Moreover, it has the following families of solutions:

(i) \[ E(t) = E_0 e^{\beta t}, \quad N(t) = 0, \]

(ii) \[ N(t) = N_0 e^{-\mu t}, \quad E(t) = 0. \]

Hence, both the \( E \) and \( N \) axes are orbits of (55). This implies that every solution \( E \) and \( N \) of (55) that starts in the first quadrant, \( E > 0 \) and \( N > 0 \), will remain there for all \( t \geq t_0 \) (which is guaranteed by the uniqueness result of the solution to (55)). Furthermore, the orbits for \( E, N \neq 0 \) can be found by solving the following equation

\[
\frac{dN}{dE} = \frac{-\mu_N N + \beta_N E N}{\beta_E E - \mu_E N E},
\]

which, after one separates variables and integrates both sides, yields

\[
\frac{N^\beta E^{\mu} e^{\beta t} E^{\mu} e^{\beta t}}{e^{\mu t} e^{\beta t} E} = k_1.
\]

Equation (56) defines a family of closed curves for \( E, N > 0 \) which are depicted in Figure 42.

![Figure 42: Orbital solutions of the predator/prey model.](image)

As shown in Figure 42, the solutions to the predator/prey model are periodic functions. The Lotka-Volterra model forms the basis of many models used today in the analysis of population dynamics. However, in its original form it has some significant problems. First, neither equilibrium point is stable (see, e.g., Figure 42). In addition, many ecologists/biologists refused to accept Volterra’s model. They cited the experiments of
G.F. Gause (1934) with two species of protozoan (one of which feeds on the other). In all experiments, the predators, Didinium, quickly destroyed the prey, Paramecium, and then died of starvation. In this case, the number of individuals of both species decays to zero and clearly does not oscillate indefinitely. Obviously the Volterra model does not take into account that bigger fish eat more and that the size varies greatly in the population. Therefore, an approach to introduce these factors into a model is to consider size-structured modeling.

31 A Size-Structured Population Model

The logistic and Lotka-Volterra models are both aggregate models. That is, they assume that all individuals are identical in characteristics and behavior (fish are all of the same size, for example). Gause’s predator/prey experiments indicate that this assumption is not very realistic.

We can attempt to model the individuals or members of a system by behavior or characteristic. This might produce a more realistic model than the aggregate model but it is also much more complicated. In 1967 Sinko and Streifer [71] balanced this trade-off by letting all individuals share some common traits, but permitted variation in size. Their formulation and its generalizations have subsequently been used widely in biological modeling [69].

Let \( u(t, x) \) be the number of individuals of size \( x \) at time \( t \). If we assume that the species has \( M \) distinct size classes \( x_1, x_2, \ldots, x_M \), the total population \( N(t) \) at time \( t \) will be given by

\[
N(t) = \sum_{i=1}^{M} u(t, x_i).
\]

This is size discrete modeling. Here, growth is a jump from one size class to the next. For growth to be continuous, we will let \( x = x(t) \) be a continuous function of \( t \). Now we cannot determine how many individuals are in a specific size class, but instead we calculate the number in an interval of size. We use \( u(t, x) \) to denote size density (in numbers per unit size) and calculate the number of individuals between size \( a \) and \( b \) at time \( t \) by

\[
N_{ab}(t) = \int_{a}^{b} u(t, \xi) d\xi.
\]

It is important to note that \( x \) is not a spatial variable and has nothing to do with the location of the individual in the medium. It actually denotes size. Since \( x(t) \) is size, the flux of \( x(t) \) is defined in terms of growth from size \( x \) to \( x + \Delta x \). Also the size density term, \( u(t, x) \), would have units of individuals/size that is very different from a location density data, which might have units of individuals/length\(^3\).
As already mentioned above, to balance an aggregate model and individual model, Sinko and Streifer grouped individuals sharing common traits together. Specifically, they make the following assumptions:

1. The growth rate, $g > 0$, of same sized individuals is the same. That is,
   \[
   \frac{dx}{dt} = g(t, x).
   \]
   The simplifying effect of this equation for growth is that the growth of all sizes of individuals is governed by this one equation. Moreover, it is assumed that $g$ is a continuous function.

2. Individuals of the same size have the same likelihood of death. In a simple version, all sizes will have the same death rate. This gives the following basic equation for “simple” death:
   \[
   \frac{dN}{dt} = -\mu N(t),
   \]
   where $\mu$ is the constant of proportionality of mortality. A more complicated model will have $\mu = \mu(x)$ so that mortality is a function of size (i.e., a large individual is more likely to die than a small individual).

3. The population is sufficiently large to be treated with a continuum model.

4. There is a “smallest” and a “largest” size ($x_0 \leq x \leq x_1$).

5. Birth (also called recruitment) rate is proportional to the population size density and is given by
   \[
   R(t) = \int_{x_0}^{x_1} k(t, \xi) u(t, \xi) d\xi,
   \]
   where $k(t, \xi)$ is the size-dependent fertility term also called the fecundity function.

Now dividing both sides of equation (57) by $\Delta x$ and letting $\Delta x$ go to zero we obtain
\[
\frac{\partial}{\partial t} u(t, x) + \frac{\partial}{\partial x} (g(t, x) u(t, x)) = -\mu(t, x) u(t, x).
\]
Equation (57) is known as the McKendrick-Von Foerster equation or the Sinko-Streifer equation [71]. The functions $g(t, x)$ and $\mu(t, x)$ correspond
respectively to the growth rate of an individual of size $x$ at time $t$ and the fraction of individuals of size $x$ dying at time $t$. To complete the description of this mathematical model requires the specification of an initial condition

$$u(0, x) = \Phi(x)$$

and a boundary condition. We assume that all births entering the population begin at the smallest size $x_0$, for simplicity. More specifically, we have:

$$\text{rate of population entering at } x_0 = \text{birth rate}$$

$$g(t, x_0)u(t, x_0) = \int_{x_0}^{x_1} k(t, \xi)u(t, \xi) \, d\xi,$$

or

$$R(t) = g(t, x)u(t, x)|_{x=x_0} = \int_{x_0}^{x_1} k(t, \xi)u(t, \xi) \, d\xi. \quad (58)$$

Here $R$ is known as the recruitment rate. We note that when the newborns enter the system, they follow the characteristic growth curves just like other individuals. The addition of (58) essentially completes the specification of the mathematical model (57). However, since $x_1$ is the maximum attainable size, we also impose the physical condition

$$g(t, x)u(t, x)|_{x=x_1} = 0.$$

### 32 Size Structure and Mosquitofish Populations

We return now to the mosquitofish populations that we introduced as motivation at the beginning of this chapter. We have discussed in previous sections the Sinko-Streifer model and methods for its solution in both forward problem and inverse problem settings. While the Sinko-Streifer model is widely (and successfully) used in the literature (see [52, 53, 68]) on biological populations, it has some rather serious shortcomings. These are readily seen in considering the mosquitofish data [63] depicted in Figure 43. In this data, we see that a pulse of population (23 July) exhibits in time both dispersion (6 August) and bifurcation (25 August). That is, a unimodal density disperses and becomes bimodal. Recalling the solution of the Sinko-Streifer equation (in particular, the initial
condition driven solution (??)), we see that a pulse propagates without dispersion or bifurcation. The initial data \( \Phi \) propagates along characteristics emanating from its region of non-zero support with amplitude increasing or decreasing in time depending on the values of \( \frac{\partial g}{\partial x} \) and \( \mu \). In fact, one can argue from the Sinko-Streifer equation itself and the method of characteristics that dispersion cannot occur unless \( \frac{\partial g}{\partial x} > 0 \); this is a condition for spreading of the characteristic curves defined by \( \frac{dx}{dt} = g \). Such an assumption is inherently unreasonable in many biological applications: in our example it is equivalent to the assumption that individual growth rates increase as one’s size increases! Indeed, it is counter intuitive that the larger one is, the faster one grows.
Figure 43: *Mosquitofish data.*
Thus, while there is no hope that the Sinko-Streifer model as developed in the previous sections can describe the mosquitofish data, one might be reluctant to abandon such a popular as well as reasonable growth model. One might instead turn to a more careful analysis of the assumptions underlying the Sinko-Streifer equation. Further investigation of the mosquitofish populations and their biological properties leads to the additional information that males and females reach different maximum sizes (30mm and 60mm, respectively). Hence, we conclude that males and females in the size range 28-30mm must grow at different rates. This immediately violates one of the underlying assumptions of the Sinko-Streifer model that individuals of the same size grow at the same rate, i.e., the assumption \( \frac{dx}{dt} = g(t, x) \) cannot be reasonable for mosquitofish (and possibly other) data. We next describe an idea first introduced in [60] and later theoretically developed in [55, 56, 57, 61] and more recently in modeling of early growth of shrimp [54, 58].

To generalize the Sinko-Streifer equation, we allow individuals of the same size to possess different individual growth rates. This can be accomplished by assuming the existence of “intrinsic” parameters \( \gamma \) (which in general we cannot observe and hence cannot use to physically distinguish individuals in the data) on which the growth rates depend. Thus we assume

\[
\frac{dx}{dt} = g(t, x; \gamma)
\]

as a parameter-dependent individual growth rate. The parameter values may range over a set of admissible parameters \( \Gamma \), and the total population is composed of subgroups, grouped together in population substructures characterized by common \( \gamma \) values. For example, if \( \Gamma = \{\gamma_1, \gamma_2\} \) (think males and females with each gender possessing a different \( \gamma \) value), and \( p_i \) is the proportion of individuals with intrinsic parameters \( \gamma_i \), then the total population density \( v(t, x) \) would be given by

\[
v(t, x) = p_1 u(t, x; \gamma_1) + p_2 u(t, x; \gamma_2),
\]

where \( u(t, x; \gamma_i) \) is the solution to the Sinko-Streifer equation using (59) with \( \gamma = \gamma_i \).

Of course, as soon as one admits the generalization, it is quite reasonable to assume multiple subclasses corresponding to a finite (or even infinite) family of \( \gamma \) values (again, think here that not all males of the same size have the same \( \gamma \) values), leading to a distribution of growth rates within the population. For \( \Gamma = \{\gamma_1, \cdots, \gamma_M\} \) with corresponding
proportions (probabilities) $p_i$, with $\sum p_i = 1$, the expression (60) for total population density generalizes to

$$v(t, x) = \sum_{i=1}^{M} p_i u(t, x; \gamma_i). \quad (61)$$

In the case of an (infinite) continuum $\Gamma$ of intrinsic parameter values, the above ideas generalize to a probability measure or distribution $P$ characterizing a distribution of the $\gamma$’s in $\Gamma$. Equation (61) becomes

$$v(t, x) = \int_{\Gamma} u(t, x; \gamma) dP(\gamma). \quad (62)$$

If the distribution $P$ is (absolutely) continuous, i.e., possesses a corresponding density $p = \frac{dP}{d\gamma}$, then one has

$$v(t, x) = \int_{\Gamma} u(t, x; \gamma) p(\gamma) d\gamma.$$

In [62] it is shown that these generalizations of Sinko-Streifer equation to include distributed growth rates do indeed allow the required dispersion and bifurcation so as to describe well the mosquitofish data depicted in Figure 43. In the exercise on distributed growth rates below the requested simulations allow the reader to demonstrate these features of the generalized Sinko-Streifer model. More recently [54, 58], the Sinko-Streifer system with growth rate distributions has been successfully used to model the variability in the early growth of shrimp. A mathematical and stochastic theoretical foundation as well as computational ideas for this formulation can be found in [54, 55, 56, 57, 61]
33 ELASTICITY

Motivation:
Morality tale; EM/Acoustic experiment—see [19]
Humbling lesson: Hazard of experiment first, and model second!!!

33.1 EM Acoustic Work

An acoustic pressure wave can be treated as a local variation in density which produces reflective interfaces. This is the underlying approach in [9, 10]. In these two efforts, the authors answer the following questions:

- Can you understand the material location by observing the EM pulse reflection?
- Can you obtain the material characteristics by reflected waves?
- Can you use an acoustic wave as a reflector for the interrogating EM pulse?

It was shown in these two works that an acoustic wave can be used as an artificial reflector. However, the subsequent experiment (see Figure 44) failed, and it turned out the acoustic pressure in the experiment was too low to produce acoustic waves (0.1 Pa in the experiment compared to $8 \times 10^8$ Pa required based on subsequent modeling). For

![Figure 44: Experiments of EM at CRSC.](image.jpg)

more information on this experiment and the “model-after-experiment” paradigm, see [19].

103
33.2 The Motivating Problem

In this problem, ideas from seismic technology and EM as well as the EM Doppler shift will be used to detect a buried explosive. An input generator similar to a thumper truck (Figure 45) will be used to strike the ground, thereby generating a vibration at the interface between air and soil, and subsequently at the interface between soil and target. A schematic of a thumper truck is shown in Figure 46 (from {http://www.geomore.com/seismic.html}).

This technology will utilize the concepts of micro-Doppler radar to the detection of shallow buried targets and builds on the concepts developed by the proposers and outlined in the text [10]. Using geophysical vibrators or “thumpers”, substantial motion will be induced in shallow layers of soil, sand and rocks. It is anticipated that the motion at
interfaces will be especially detected by Doppler spreading of the returned electromagnetic signal. Conductive interfaces will provide a very pronounced Doppler spectrum while non-conductive interfaces (rock versus explosive in a plastic container) will be discriminated using Doppler spectrum differences. There is of course, a long history of the use of ground penetrating radar to detect subsurface structures but soil variations have prohibited the use of ground penetrating radar in certain regions of the world. Additionally, ground penetrating radar has not had the ability to discriminate various structures of comparable size and shape. The combination of electromagnetic field and vibration in an elastic medium is expected to remedy this discrimination issue. Our belief, based on discussions with seismic experts, is that purely seismic methods are not reliable for shallow targets when used alone (also, waiting for acoustic wave returns and their post processing as in standard seismic exploration would be prohibitive from a time perspective). On the other hand, lateral electromagnetic wave propagation has promise as a shallow buried target modality; however, these waveforms may be small and pose signal-to-noise difficulty: there has been little experience with this wave configuration for detection. In a similar manner, target electromagnetic resonance responses can be used as a detector, but signal to noise issues will again pose challenges. As a proof-of-concept effort, we will first pose this problem in a 1-dimensional geometric setting as illustrated in Figure 47.

![Figure 47: Geometry of buried body target problem.](image)

In this project, the following aspects need to be explored:

- EM Doppler shifts in reflections from vibrating bodies;
- Rigid body dynamics (rigid body vibration in an elastic medium);
- Propagation of elastic waves in a dissipative medium (with possible viscoelastic properties) and their ability to produce detectable vibrations of rigid bodies.

While there are a number of interesting applications [13, 14, 15, 16, 21, 24] where modeling in elastic/viscoelastic materials is fundamental, we will focus here on the problem outlined above.
33.3 Notation

In these notes, bold letters are used to denote vectors unless otherwise indicated. For convenience of presentation, we may occasionally use Einstein notation (or index notation), where the convention is as follows: the repetition of an index in a term will denote a summation with respect to that index over its range. For example,

<table>
<thead>
<tr>
<th>Einstein notation</th>
<th>Regular notation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda \varepsilon_{kk}$</td>
<td>$\lambda \sum_k \varepsilon_{kk}$</td>
</tr>
<tr>
<td>$C_{ijkl}\varepsilon_{kl}$</td>
<td>$\sum_k \sum_l C_{ijkl}\varepsilon_{kl}$</td>
</tr>
<tr>
<td>$\delta_{ij}dx_idx_j$</td>
<td>$\sum_i \sum_j \delta_{ij}dx_idx_j$</td>
</tr>
<tr>
<td>$\frac{\partial u_k}{\partial x_i} \frac{\partial u_k}{\partial x_j}$</td>
<td>$\sum_k \frac{\partial u_k}{\partial x_i} \frac{\partial u_k}{\partial x_j}$</td>
</tr>
</tbody>
</table>

In the Cartesian coordinate system we may denote coordinate axes by $x$, $y$, and $z$ or by $x_1$, $x_2$ and $x_3$, depending on the ease of presentation.

34 Elasticity

There is a range of modeling in the mechanical problems of interest. These are

point dynamics → rigid body dynamics → continuum mechanics → fluid mechanics.

Elasticity is the physical property of a material when it deforms under stress (e.g., external forces), but returns to its original shape when the stress is removed.

34.1 Introduction

Continuum: assumes the substance of the body is distributed throughout (and completely fills) the space it occupies, and allows the approximation of physical quantities, such as energy and momentum, at the infinitesimal limit. Hence, differential equations can be employed in solving problems in continuum mechanics, in which we can study the following

- Fundamental physical laws
  - Conservation of mass (the equation of continuity)
  - Conservation of momentum (the equations of motion)
  - Conservation of energy (the first law of thermodynamics).
• Constitutive equations: this is specific to the materials being investigated. For example, we often have stress = Φ(strain) or Φ(strain, strain rate).

34.1.1 Kinematics: Deformation and Motion

An external force results in a displacement. The displacement of a body has two components:

• A rigid-body displacement: the relative displacement between particles is zero, i.e., the shape and size of the body does not change.

• A deformation: there is a relative displacement between particles, i.e, the shape and/or size are changed.
  
  – Finite strain theory: deals with deformations in which both rotations and strains are arbitrarily large. This is commonly the case with elastomers, plastically-deforming materials and other fluids and biological soft tissue. These frequently undergo large deformations and may also require viscoelastic ideas and formulations.
  
  – Infinitesimal strain theory: deals with infinitesimal deformations of a continuum body. The infinitesimal strain theory is used in the analysis of deformations of materials exhibiting classical elastic behaviour, such as materials found in mechanical and civil engineering applications, e.g., concrete and steel which typically undergo small deformations.

Here in these notes, we will focus on deformations, which is the theory of elasticity. When analyzing the deformation or motion of solids, or the flow of fluids, it is necessary to describe the sequence or evolution of configurations throughout time. One description for motion is made in terms of the material or fixed referential coordinates, and is called a material description or the Lagrangian description. The other description for motion is made in terms of the spatial or current coordinates, called a spatial description or Eulerian description. An intuitive comparison of these two descriptions would be in the Eulerian description one places the coordinate or reference system for motion of an object on the object as it moves through a moving fluid (e.g., on a boat in a river) while in the Lagrangian description one observes and describes the motion of the object from a fixed vantage point (e.g., motion of the boat from a fixed point on a bridge over the river or on the side of the river.).

• Lagrangian Description: an observer standing in the referential frame observes the changes in the position and physical properties as the material body moves in space as time progresses. In other words, this formulation focuses on individual particles as they move through space and time. This description is normally used
in solid mechanics. In the Lagrangian description, the motion of a continuum body is expressed by the mapping function $h$:

$$x = h(X, t),$$  \hspace{1cm} (63)

which is a mapping from initial (undeformed/material) configuration $\Omega_0$ to the present (deformed/spatial) configuration $\Omega_t$. Hence, in a Lagrangian coordinate system the velocity of a particle at $X$ at time $t$ is given by

$$V(X, t) = \frac{\partial x}{\partial t} = \frac{\partial h(X, t)}{\partial t},$$

and the total derivative (or material derivative) of a function $(X, t)$, which is denoted by a dot or the symbol $\frac{D}{Dt}$, is just the partial derivative of with respect to $t$, that is,

$$\frac{D}{Dt} (X, t) = \frac{\partial}{\partial t} (X, t).$$

- **Eulerian Description**: focuses on the current configuration $\Omega_t$, giving attention to what is occurring at a moving material point in space as time progresses. That is, the coordinate system is relative to a moving point in the body and hence is a moving coordinate system. This approach is conveniently applied in the study of fluid mechanics. Mathematically, the motion of a continuum using the Eulerian description is expressed by the mapping function

$$X = h^{-1}(x, t),$$  \hspace{1cm} (64)

which provides a tracing of the particle which now occupies the position $x$ in the current configuration $\Omega_t$ to its original position $X$ in the initial configuration $\Omega_0$. The velocity of a particle at $x$ at time $t$ in Eulerian coordinate system is

$$v(x, t) = V(h^{-1}(x, t), t).$$

Hence, in Eulerian coordinate system the total derivative (or material derivative) of a function $(x, t)$ is given by

$$\frac{D}{Dt} (x, t) = \frac{\partial}{\partial t} (x, t) + \sum_{i=1}^{3} v_i \frac{\partial}{\partial x_i} (x, t) = \frac{\partial}{\partial t} (x, t) + v(x, t) \cdot \nabla (x, t).$$

**Remark 34.1** The following are the names used in the literature to refer to Lagrangian and Eulerian configurations.

- **Lagrangian**: initial/referential, material, undeformed, fixed.
- **Eulerian**: current/present, space, deformed, moving.
34.1.2 Displacement and Strain

A particle \( P \) located originally at a place with coordinate \( \mathbf{X} = (X_1, X_2, X_3) \) is moved to a place \( P' \) with coordinate \( \mathbf{x} = (x_1, x_2, x_3) \) when the body moves and deforms. Then the vector \( \mathbf{PP}' \), is called the displacement or deformation vector of the particle. The displacement vector is, clearly,

\[
\mathbf{x} - \mathbf{X}
\]  

(65)

Let the variable \( \mathbf{X} = (X_1, X_2, X_3) \) identify a particle in the original configuration of the body, and \( \mathbf{x} = (x_1, x_2, x_3) \) be the coordinates of that particle when the body is deformed. Then the deformation of a body is known if \( x_1, x_2 \) and \( x_3 \) are known functions of \( X_1, X_2, X_3 \):

\[
x_i = x_i(X_1, X_2, X_3), \quad i = 1, 2, 3.
\]  

(66)

Then the displacement (which will subsequently called the Lagrangian displacement) of the particle relative to \( \mathbf{X} \) is given by

\[
\mathbf{U}(\mathbf{X}) = \mathbf{x}(\mathbf{X}) - \mathbf{X}.
\]  

(67)

If we assume the transformation has a unique inverse, then we have

\[
X_i = X_i(x_1, x_2, x_3), \quad i = 1, 2, 3.
\]  

(68)

for every particle in the body. Then the (Eulerian) displacement of the particle relative to \( \mathbf{x} \) is given by

\[
\mathbf{u}(\mathbf{x}) = \mathbf{x} - \mathbf{X}(\mathbf{x}).
\]  

(69)

A rigid body expresses no internal stress. Thus the displacements themselves are not directly related to stress. To relate deformation with stress, we must consider the stretching and distortion of the body. For this purpose, it is sufficient if we know the change of distance between any arbitrary pair of points.

Consider an infinitesimal line segment connecting the point \( P(X_1, X_2, X_3) \) to a neighboring point \( Q(X_1 + dX_1, X_2 + dX_2, X_3 + dX_3) \) (see Figure 48). The square of the lengths of \( PQ \) in the original configuration is given by

\[
|d\mathbf{X}|^2 = (d\mathbf{X})^T d\mathbf{X} = (dX_1)^2 + (dX_2)^2 + (dX_3)^2.
\]

When \( P \) and \( Q \) are deformed to the points \( P'(x_1, x_2, x_3) \) and \( Q'(x_1 + dx_1, x_2 + dx_2, x_3 + dx_3) \), respectively, the square of length of \( P'Q' \) is

\[
|d\mathbf{x}|^2 = (d\mathbf{x})^T d\mathbf{x} = (dx_1)^2 + (dx_2)^2 + (dx_3)^2.
\]

Definition 34.1 The configuration gradient is defined by

\[
A = \frac{d\mathbf{x}}{d\mathbf{X}} = \begin{bmatrix}
\frac{\partial x_1}{\partial X_1} & \frac{\partial x_1}{\partial X_2} & \frac{\partial x_1}{\partial X_3} \\
\frac{\partial x_2}{\partial X_1} & \frac{\partial x_2}{\partial X_2} & \frac{\partial x_2}{\partial X_3} \\
\frac{\partial x_3}{\partial X_1} & \frac{\partial x_3}{\partial X_2} & \frac{\partial x_3}{\partial X_3}
\end{bmatrix}.
\]  

(70)
Definition 34.2 The right Cauchy-Green configuration tensor is defined by

\[ D_R = A^T A = \begin{pmatrix} \frac{\partial x_k}{\partial X_i} \partial x_k \end{pmatrix}. \]

Invariants of \( D_R \) are often used in the expressions for strain energy density functions (to be discussed later). The most common used invariants are

\[ I_1 = \text{tr}(D_R) = \lambda_1^2 + \lambda_2^2 + \lambda_3^2, \]
\[ I_2 = \frac{1}{2} [\text{tr}(D_R^2) - (\text{tr}(D_R))^2] = \lambda_1^2 \lambda_2^2 + \lambda_2^2 \lambda_3^2 + \lambda_3^2 \lambda_1^2, \]
\[ I_3 = \det(D_R) = \lambda_1^2 \lambda_2^2 \lambda_3^2, \]

where \( \lambda_i, i = 1, 2, 3 \) are the eigenvalues of \( A \), and also known as principal stretches as will be discussed later. Hence, the invariants are defined to be the coefficients of the characteristic equation of \( D_R \).

The left Cauchy-Green configuration tensor is defined by

\[ D_L = AA^T = \left( \begin{pmatrix} \frac{\partial x_i}{\partial X_k} \partial x_k \end{pmatrix} \right). \]

The invariants for \( D_L \) are defined as the same as those for \( D_R \).

By the definition of configuration gradient, we have (recall (63)) \( \text{d}x = AdX \) and

\[ |\text{d}x|^2 - |\text{d}X|^2 = (\text{d}x)^T dX - (\text{d}X)^T dX = (\text{d}X)^T A^T AdX - (\text{d}X)^T dX = (\text{d}X)^T (A^T A - I) dX \]

If strain (as defined below) satisfies \( A^T A - I = 0 \), then we say the object is undeformed. Otherwise, it is deformed.
The Lagrangian finite strain tensor

Lagrangian finite strain tensor $E$ is defined by

$$E = \frac{1}{2}(A^T A - I).$$

(73)

Now we explore the relationship between the displacement and strain. By (67) and (70) we have the deformation gradient given by

$$\nabla U = \begin{bmatrix} \frac{\partial x_1}{\partial X_1} - 1 & \frac{\partial x_1}{\partial X_2} & \frac{\partial x_1}{\partial X_3} \\ \frac{\partial x_2}{\partial X_1} & \frac{\partial x_2}{\partial X_2} - 1 & \frac{\partial x_2}{\partial X_3} \\ \frac{\partial x_3}{\partial X_1} & \frac{\partial x_3}{\partial X_2} & \frac{\partial x_3}{\partial X_3} - 1 \end{bmatrix} = A - I$$

(74)

or

$$\frac{\partial U_i}{\partial X_j} = \frac{\partial x_i}{\partial X_j} - \delta_{ij}, \; j = 1, 2, 3, \; i = 1, 2, 3.$$  

(75)

Thus, because $A = \nabla U + I$, the relationship between Lagrangian finite strain (73) and displacement is given by

$$E_{ij} = \frac{1}{2} \left[ \frac{\partial U_i}{\partial X_j} + \frac{\partial U_j}{\partial X_i} + \frac{\partial U_k}{\partial X_i} \frac{\partial U_k}{\partial X_j} \right].$$

(76)

The Eulerian finite strain tensor

Similarly, using $dX = A^{-1}dx$, we find

$$|dx|^2 - |dX|^2 = (dx)^T dx - (dx)^T (A^{-1})^T A^{-1} dx$$

(77)

$$= (dx)^T (I - (A^{-1})^T A^{-1}) dx$$

and the Eulerian finite strain tensor $e$ is defined by

$$e = \frac{1}{2} (I - (A^{-1})^T A^{-1}).$$

(78)

Now we explore the relationship between the displacement and strain. By (69) and (70) we have

$$\nabla u = \begin{bmatrix} 1 - \frac{\partial X_1}{\partial x_1} & -\frac{\partial X_1}{\partial x_2} & -\frac{\partial X_1}{\partial x_3} \\ -\frac{\partial X_2}{\partial x_1} & 1 - \frac{\partial X_2}{\partial x_2} & -\frac{\partial X_2}{\partial x_3} \\ -\frac{\partial X_3}{\partial x_1} & -\frac{\partial X_3}{\partial x_2} & 1 - \frac{\partial X_3}{\partial x_3} \end{bmatrix} = I - A^{-1}$$

(79)

or

$$\frac{\partial u_i}{\partial x_j} = \delta_{ij} - \frac{\partial X_i}{\partial x_j}, \; j = 1, 2, 3, \; i = 1, 2, 3.$$  

(80)
Thus, the relationship between Eulerian finite strain and displacement is given by
\[
e_{ij} = \frac{1}{2} \left[ \frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} - \frac{\partial u_k}{\partial x_i} \frac{\partial u_k}{\partial x_j} \right].
\] (81)

**Infinitesimal strain theory**: If the components of displacement \( u_i \) are such that their first derivatives are so small such that the squares and the products of the partial derivatives of \( u_i \) are negligible compared with the first-order terms, then \( e_{ij} \) reduces to Cauchy’s infinitesimal strain tensor
\[
\varepsilon_{ij} = \frac{1}{2} \left[ \frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right],
\] (82)
which implies that the infinitesimal strain tensor is symmetric, i.e., \( \varepsilon_{ij} = \varepsilon_{ji} \).

Similarly, the Lagrangian finite strain \( E_{ij} \) in the infinitesimal strain theory reduces to
\[
\mathcal{E}_{ij} = \frac{1}{2} \left[ \frac{\partial U_i}{\partial X_j} + \frac{\partial U_j}{\partial X_i} \right].
\] (83)

### 34.1.3 Stress

Stress is a measure of the average amount of force exerted per unit area (\( N/m^2 \) or Pa), and it is a reaction to external forces. Stress was introduced into the theory of elasticity by Cauchy around 1822.

**Definition 34.3** *The stress vector (traction) is defined by*

\[
T^{(n)} = \frac{d\mathbf{F}}{d\Gamma},
\] (84)

*where the superscript \((n)\) is introduced to denote the direction of the normal vector \( \mathbf{n} \) of the surface \( \Gamma \) and \( \mathbf{F} \) is the force on the surface.*

Consider a little cube in the body as shown in Figure 49 (left). Let the surface of the cube normal (perpendicular) to the axis \( z \) be donated by \( \nabla \Gamma_z \). Let the stress vector that acts on the surface \( \nabla \Gamma_z \) be \( T^{(e_3)} \), where \( e_3 = (0, 0, 1) \). Resolve \( T^{(e_3)} \) into three components in the direction of the coordinate axes and denote them by \( S_{xx}, S_{yy} \) and \( S_{zz} \). Similarly we may consider surface \( \nabla \Gamma_x \) and \( \nabla \Gamma_y \) perpendicular to \( x \) and \( y \), the stress vectors acting on them, and their components in the \( x, y \) and \( z \) directions. The components \( S_{xx}, S_{yy} \) and \( S_{zz} \) are normal stresses, and \( S_{xy}, S_{xz}, S_{yx}, S_{yz}, S_{zx} \) and \( S_{zy} \) are shear stresses. *A stress component is positive if it acts in the positive direction of the coordinate axes.* We remark that the notation \( S_{\text{face}, \text{direction}} \) is consistently used in elasticity theory.
Definition 34.4  The Cauchy stress tensor is defined by

\[
\mathbf{S} = \begin{bmatrix} T^{(e_1)} \\ T^{(e_2)} \\ T^{(e_3)} \end{bmatrix} = \begin{bmatrix} S_{xx} & S_{xy} & S_{xz} \\ S_{yx} & S_{yy} & S_{yz} \\ S_{zx} & S_{zy} & S_{zz} \end{bmatrix},
\]

where \( e_1 = (1, 0, 0), e_2 = (0, 1, 0) \) and \( e_3 = (0, 0, 1) \).

Theorem 34.2  Let \( \mathbf{T}^{(n)} \) be the stress vector acting on \( d\Gamma \) whose outer normal vector is \( \mathbf{n} \), which is illustrated in Figure 50. Cauchy’s formula expresses \( \mathbf{T}^{(n)} \) as a function

Figure 49: Notations of stress components.

Figure 50: Stress vector acting on a plane with normal \( \mathbf{n} \).
of the stress vectors on the planes perpendicular to the coordinate axes, i.e., in terms of the components of the Cauchy stress tensor. This formula (page 69 in [18]) asserts that

\[
\begin{align*}
T_x^{(n)} &= S_{xx} n_x + S_{yx} n_y + S_{zx} n_z, \\
T_y^{(n)} &= S_{xy} n_x + S_{yy} n_y + S_{zy} n_z, \\
T_z^{(n)} &= S_{xz} n_x + S_{yz} n_y + S_{zz} n_z.
\end{align*}
\] (86)

Here \( T^{(n)} = (T_x^{(n)}, T_y^{(n)}, T_z^{(n)})^T \), and \( n = (n_x, n_y, n_z)^T \). Cauchy’s formula (86) can be written concisely as follows:

\[
T^{(n)} = S^{T} n,
\]

where \( S \) is the Cauchy stress tensor defined in (85).

**Remark 34.3** A great diversity in notations for stress components exists in the literature. The most widely used notation in American literature (according to Fung) is, in reference to a system of rectangular Cartesian coordinates \( x, y, z \),

\[
\begin{bmatrix}
\sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\
\sigma_{yx} & \sigma_{yy} & \sigma_{yz} \\
\sigma_{zx} & \sigma_{zy} & \sigma_{zz}
\end{bmatrix}, \quad \text{or} \quad \begin{bmatrix}
\sigma_{11} & \sigma_{12} & \sigma_{13} \\
\sigma_{21} & \sigma_{22} & \sigma_{23} \\
\sigma_{31} & \sigma_{32} & \sigma_{33}
\end{bmatrix}.
\]

In the engineering literature, rectangular Cartesian coordinates are usually denoted by \( x_1, x_2, x_3 \) instead of \( x, y, z \). Hence, the notation for stress components is also commonly written as

\[
\begin{bmatrix}
\tau_{xx} & \tau_{xy} & \tau_{xz} \\
\tau_{yx} & \tau_{yy} & \tau_{yz} \\
\tau_{zx} & \tau_{zy} & \tau_{zz}
\end{bmatrix}.
\]

We use here \( S \) instead of \( \sigma \) or \( \tau \) because in EM theory \( \sigma \) often denotes the conductivity of material and \( \tau \) is the relaxation time of a material. In these notes, for presentation convenience we may also use \( \sigma_{ij} \) for stress components.

**Remark 34.4** The Cauchy stress tensor is not the only measure of stress that is used in practice. Other measures of stress include the first and second Piola-Kirchhoff stress tensors, the Biot stress tensor, and the Kirchhoff stress tensor. Their differences are illustrated as follows:

- **Cauchy stress tensors**: relates forces in the present (deformed/spatial) configuration to areas in the present configuration. Hence, sometimes Cauchy stress is also called true stress. In addition, the Cauchy stress tensor is symmetric.

- **Piola-Kirchhoff stress tensors** are used to express the stress relative to the fixed reference (undeformed/material) configuration.
First Piola-Kirchhoff stress tensor: relates forces in the present configuration with areas in the reference configuration. This stress tensor is usually not symmetric, and its relationship with the Cauchy stress tensor is expressed below in (103).

Second Piola-Kirchhoff stress tensor: relates forces in the reference configuration to areas in the reference configuration.

Please note that for infinitesimal deformations, the Cauchy and Piola-Kirchhoff tensors are identical. For more information on these stress tensors, please refer to [http://en.wikipedia.org/wiki/Stress](http://en.wikipedia.org/wiki/Stress).

### 34.2 Derivation of Equations of Motion of a Continuum

Before deriving the equations of motion of a continuum, we first discuss forces. There are two types of external forces acting on material bodies in the mechanics of continuum media:

- **Body forces** \((N/m^3)\), acting on elements of volume of body.
  - For example, gravitational force and EM force.

- **Surface forces** \((N/m^2)\), or stress, acting on surface elements.
  - For example, aerodynamics pressure acting on a body, stress between one part of a body on another, etc.

Then the total force \(\mathbf{F}\) acting upon the material occupying the region \(\Omega\) interior to a closed surface \(\Gamma\) is

\[
\mathbf{F} = \oint_{\Gamma} \mathbf{T}^{(n)} d\Gamma + \int_{\Omega} \mathbf{f} d\Omega, \tag{87}
\]

where \(\mathbf{f} = (f_x, f_y, f_z)^T\) is the body force, and \(\mathbf{T}^{(n)}\) is the stress vector acting on \(d\Gamma\) whose outer normal vector is \(\mathbf{n}\).

The expression (87) is a universal force balance statement independent of any particular coordinate system being used. Of course, with either the Eulerian or Lagrangian formulation, the stresses and forces must be expressed in terms of the appropriate coordinate system.

In this section, we will use an integral approach to derive the equation of continuity and the equation of motion of a continuum, first in the Eulerian (or moving) coordinate system and then in the Lagrangian coordinate system. In the following, we will sometimes use \(\Omega\) to denote \(\Omega_t\) for ease in the presentation. But either will refer to a volume element that is time dependent.
34.2.1 The Material Derivative of a Volume Integral

To carry out our derivations, we need a calculus for interchanging integration and differentiation when both the limits of the integration and the integrand depend on the differentiation variable. Let \( \Phi(t) \) be a volume integral of a continuously differential function \( \phi(x, y, z, t) \) defined over a spatial domain \( \Omega_t \) occupied by a given set of material particles at time \( t \):

\[
\Phi(t) = \int_{\Omega_t} \phi(x, y, z, t) d\Omega.
\]  

Then the rate of change of \( \Phi(t) \) with respect to \( t \) is given by (we suppress the multiple integral notation here and below when it is clearly understood that the integral is a volume or surface integral)

\[
\frac{d\Phi}{dt} = \int_{\Omega_t} \frac{\partial \phi}{\partial t} d\Omega + \int_{\Gamma_t} (\phi v_x n_x + \phi v_y n_y + \phi v_z n_z) d\Gamma,
\]  

where on the boundary \( \Gamma = \Gamma_t \) of \( \Omega_t \), \( v = v(t) \) is the velocity \( v(t) = (\frac{dx}{dt}, \frac{dy}{dt}, \frac{dz}{dt}) \). This can be written concisely as

\[
\frac{d\Phi}{dt} = \int_{\Omega_t} \frac{\partial \phi}{\partial t} d\Omega + \int_{\Gamma_t} \phi \cdot n d\Gamma.
\]

The 1st term of the right side corresponds to rate of change in a fixed volume, and the 2nd term corresponds to the convective transfer through the surface. By Gauss’s theorem, the above equality can also be written as

\[
\frac{d\Phi}{dt} = \int_{\Omega_t} \left( \frac{\partial \phi}{\partial t} + \frac{\partial \phi v_x}{\partial x} + \frac{\partial \phi v_y}{\partial y} + \frac{\partial \phi v_z}{\partial z} \right) d\Omega,
\]  

or written concisely as

\[
\frac{d\Phi}{dt} = \int_{\Omega_t} \left( \frac{\partial \phi}{\partial t} + \nabla \cdot (\phi v) \right) d\Omega.
\]

This rate, called the material derivative of \( \Phi \), is defined for a given set of material particles in a moving volume. Note that when \( \Omega_t = \Omega_0 \) for all \( t \) (i.e., the boundary \( \Gamma \) is not moving so that \( v = 0 \)), this becomes simply

\[
\frac{d}{dt} \int_{\Omega_0} \phi(x, y, z, t) d\Omega = \int_{\Omega_0} \frac{\partial \phi}{\partial t} d\Omega.
\]  

34.2.2 The Equations of Continuity

We next derive the equations of continuity for an arbitrary mass of particles that may be changing in time. The mass contained in a domain \( \Omega_t \) at time \( t \) is

\[
m(t) = \int_{\Omega_t} \rho(x, y, z, t) d\Omega.
\]
Conservation of mass requires that $\frac{dm}{dt} = 0$ and thus we have from (90)

$$\frac{dm}{dt} = \int_{\Omega_t} \left[ \frac{\partial \rho}{\partial t} + \frac{\partial \rho v_x}{\partial x} + \frac{\partial \rho v_y}{\partial y} + \frac{\partial \rho v_z}{\partial z} \right] d\Omega. \tag{93}$$

Hence, we have

$$\int_{\Omega_t} \left[ \frac{\partial \rho}{\partial t} + \frac{\partial \rho v_x}{\partial x} + \frac{\partial \rho v_y}{\partial y} + \frac{\partial \rho v_z}{\partial z} \right] d\Omega = 0.$$

Since the above equality holds for an arbitrary domain $\Omega_t$, we obtain the equation of continuity

$$\frac{\partial \rho}{\partial t} + \frac{\partial \rho v_x}{\partial x} + \frac{\partial \rho v_y}{\partial y} + \frac{\partial \rho v_z}{\partial z} = 0, \tag{94}$$

which can be written concisely as

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

### 34.2.3 Reynolds Transport Theorem

In this subsection, we will use the material derivative (90) as well as the equation of continuity (94) to derive the famous Reynolds’ transport theorem. By (90) we find that

$$\frac{d}{dt} \int_{\Omega_t} \rho v_z d\Omega = \int_{\Omega_t} \left( \frac{\partial (\rho v_z)}{\partial t} + \frac{\partial \rho v_z v_x}{\partial x} + \frac{\partial \rho v_z v_y}{\partial y} + \frac{\partial \rho v_z v_z}{\partial z} \right) d\Omega.\tag{95}$$

Then by (94), we find that the integrand of the right side of the above equation is equal to

$$\frac{\partial \rho v_z}{\partial t} + \rho \frac{\partial v_z}{\partial t} + v_z \left( \frac{\partial \rho v_z}{\partial x} + \frac{\partial \rho v_z}{\partial y} + \frac{\partial \rho v_z}{\partial z} \right) + \rho v_x \frac{\partial v_z}{\partial x} + \rho v_y \frac{\partial v_z}{\partial y} + \rho v_z \frac{\partial v_z}{\partial z}$$

$$= v_z \left( \frac{\partial \rho}{\partial t} + \frac{\partial \rho v_x}{\partial x} + \frac{\partial \rho v_y}{\partial y} + \frac{\partial \rho v_z}{\partial z} \right) + \rho v_x \frac{\partial v_z}{\partial x} + \rho v_y \frac{\partial v_z}{\partial y} + \rho v_z \frac{\partial v_z}{\partial z}$$

$$= \rho \left( \frac{\partial v_z}{\partial t} + v_x \frac{\partial v_z}{\partial x} + v_y \frac{\partial v_z}{\partial y} + v_z \frac{\partial v_z}{\partial z} \right) \text{ according to equation of continuity.}$$

Hence, we have

$$\frac{d}{dt} \int_{\Omega_t} \rho v_z d\Omega = \int_{\Omega_t} \rho \left( \frac{\partial v_z}{\partial t} + v_x \frac{\partial v_z}{\partial x} + v_y \frac{\partial v_z}{\partial y} + v_z \frac{\partial v_z}{\partial z} \right) d\Omega.\tag{95}$$

Equation (95) is the famous Reynold’s transport theorem, which is usually written concisely as

$$\frac{d}{dt} \int_{\Omega_t} \rho v_z d\Omega = \int_{\Omega_t} \rho \frac{Dv_z}{Dt} d\Omega,$$
where \( \frac{Dv_z}{Dt} \) is the total derivative of \( v_z \), and is given by

\[
\frac{Dv_z}{Dt}(x, y, z, t) = \frac{\partial v_z}{\partial t} + v_x \frac{\partial v_z}{\partial x} + v_y \frac{\partial v_z}{\partial y} + v_z \frac{\partial v_z}{\partial z}.
\]

We note that the above is independent of any coordinate system and depends only on the rules of calculus and the assumptions of continuity of mass in a time dependent volume of particles.

### 34.2.4 The Eulerian Equations of Motion of a Continuum

We are now ready to use the above rules of calculus and the continuity of mass as embodied in the Reynolds theorem to derive the equations of motion in an Eulerian coordinate system. Throughout we have \( \Omega = \Omega_t \) and \( \Gamma = \Gamma_t \) (though we suppress the subscripts) and we assume the coordinate system \((x, y, z)\) is now moving (changing with the volume element) with a velocity \( v(t) = (\frac{dx}{dt}, \frac{dy}{dt}, \frac{dz}{dt})\) of the deformation of the material.

The resultant force \( F_z \) in the \( z \)-direction on an arbitrary volume \( \Omega \) is

\[
F_z = \int_\Omega T_z^{(n)} d\Omega + \int_\Omega f_z d\Omega. \tag{96}
\]

By Cauchy’s formula (86) and Gauss’s theorem we have

\[
\int_\Gamma T_z^{(n)} d\Gamma = \int_\Gamma (S_{xz} n_x + S_{yz} n_y + S_{zz} n_z) d\Gamma = \int_\Omega \left( \frac{\partial S_{xz}}{\partial x} + \frac{\partial S_{yz}}{\partial y} + \frac{\partial S_{zz}}{\partial z} \right) d\Omega.
\]

Hence, by the above equality and (96) we obtain that

\[
F_z = \int_\Omega \left( \frac{\partial S_{xz}}{\partial x} + \frac{\partial S_{yz}}{\partial y} + \frac{\partial S_{zz}}{\partial z} + f_z \right) d\Omega. \tag{97}
\]

Newton’s law states that

\[
\frac{d}{dt} \int_\Omega \rho v_z d\Omega = \int_\Omega \left( \frac{\partial S_{xz}}{\partial x} + \frac{\partial S_{yz}}{\partial y} + \frac{\partial S_{zz}}{\partial z} + f_z \right) d\Omega. \tag{98}
\]

Hence, by Reynold’s transport theorem we have that

\[
\int_\Omega \rho \left( \frac{\partial v_z}{\partial t} + v_x \frac{\partial v_z}{\partial x} + v_y \frac{\partial v_z}{\partial y} + v_z \frac{\partial v_z}{\partial z} \right) d\Omega = \int_\Omega \left( \frac{\partial S_{xz}}{\partial x} + \frac{\partial S_{yz}}{\partial y} + \frac{\partial S_{zz}}{\partial z} + f_z \right) d\Omega.
\]

Note that because the above equality holds for an arbitrary domain \( \Omega \), the integrands on both sides must be equal. Thus, we have

\[
\rho \left( \frac{\partial v_z}{\partial t} + v_x \frac{\partial v_z}{\partial x} + v_y \frac{\partial v_z}{\partial y} + v_z \frac{\partial v_z}{\partial z} \right) = \frac{\partial S_{xz}}{\partial x} + \frac{\partial S_{yz}}{\partial y} + \frac{\partial S_{zz}}{\partial z} + f_z, \tag{99}
\]

118
or written concisely as
\[
\rho \frac{Dv_z}{Dt} = \nabla \cdot S_{\bullet z} + f_z,
\]
which is the equation of motion of a continuum in the z-direction. The entire set for the equations of motion of a continuum in Eulerian coordinate system is given as follows:

\[
\rho \left( \frac{\partial v_x}{\partial t} + v_x \frac{\partial v_x}{\partial x} + v_y \frac{\partial v_x}{\partial y} + v_z \frac{\partial v_x}{\partial z} \right) = \frac{\partial S_{xx}}{\partial x} + \frac{\partial S_{yx}}{\partial y} + \frac{\partial S_{zx}}{\partial z} + f_x
\]

\[
\rho \left( \frac{\partial v_y}{\partial t} + v_x \frac{\partial v_y}{\partial x} + v_y \frac{\partial v_y}{\partial y} + v_z \frac{\partial v_y}{\partial z} \right) = \frac{\partial S_{xy}}{\partial x} + \frac{\partial S_{yy}}{\partial y} + \frac{\partial S_{zy}}{\partial z} + f_y
\]

\[
\rho \left( \frac{\partial v_z}{\partial t} + v_x \frac{\partial v_z}{\partial x} + v_y \frac{\partial v_z}{\partial y} + v_z \frac{\partial v_z}{\partial z} \right) = \frac{\partial S_{xz}}{\partial x} + \frac{\partial S_{yz}}{\partial y} + \frac{\partial S_{zz}}{\partial z} + f_z.
\]

We note that (100) is also called Cauchy’s equation of motion or Cauchy’s momentum equation in some literature. Equation (100) can be written in vector form as

\[
\rho \left( \frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla)\mathbf{v} \right) = \nabla \cdot \mathbf{S}^T + \mathbf{f},
\]

where \( \mathbf{S} \) is the Cauchy stress tensor defined in (85). It is often desirable to express these equations of motion in terms of displacements \( \mathbf{u} \). We find (because the Eulerian velocity is given in terms of the displacement by \( \mathbf{v} = \frac{\partial \mathbf{u}}{\partial t} \))

\[
\rho \left( \frac{\partial^2 \mathbf{u}}{\partial t^2} + (\frac{\partial \mathbf{u}}{\partial t} \cdot \nabla)\frac{\partial \mathbf{u}}{\partial t} \right) = \nabla \cdot \mathbf{S}^T + \mathbf{f}.
\]

### 34.2.5 The Lagrangian Equations of Motion of a Continuum

Next we will rewrite (100) in terms of Lagrangian description, that is, we will derive an equation of motion in Lagrangian coordinate system \((O - XYZ\) coordinate system). Let \( \Gamma_0 \) denote the boundary of \( \Omega_0 \) in the initial (undeformed/material) configuration, and \( \mathbf{n}_0 \) be the outer normal vector on \( \Gamma_0 \). By Nanson’s formula [23] we have

\[
\mathbf{n} d\Gamma = |A| (A^{-1})^T \mathbf{n}_0 d\Gamma_0,
\]

where \( \mathbf{n}_0 = (n_{0X}, n_{0Y}, n_{0Z})^T \), \( A \) is the configuration gradient defined by (70), and \( |A| \) is the determinant of \( A \). Multiplying both sides of (102) by \( \mathbf{S}^T \) yields that

\[
\mathbf{S}^T \mathbf{n} d\Gamma = |A| \mathbf{S}^T (A^{-1})^T \mathbf{n}_0 d\Gamma_0,
\]

where \( \mathbf{S} \) is Cauchy stress tensor defined by (85). Let

\[
\mathbf{P} = |A| \mathbf{S}^T (A^{-1})^T,
\]

119
which is the first Piola-Kirchhoff stress tensor. Then we have

\[
S^T \mathbf{nd} \Gamma = \mathbf{Pn_0} d\Gamma_0. 
\]

Let \( \mathbf{f_0} \) be the external body force acting on \( \Omega_0 \) (\( \mathbf{f_0} = |A| \mathbf{f} \)), let \( \rho_0(X, Y, Z, t) \) be the material density in the Lagrangian coordinate system (conservation of mass implies that \( \rho_0 = |A| \rho \)), and \( V(X, Y, Z, t) \) be the velocity in the Lagrangian coordinate system. Then we can rewrite the resultant force in the \( z \)-direction in the Eulerian coordinate system as the resultant force in the \( Z \) direction in the Lagrangian coordinate system, which is

\[
\mathbf{F_0Z} = \int_{\Gamma_0} (P_{ZX}n_{0X} + P_{ZY}n_{0Y} + P_{ZZ}n_{0Z}) d\Gamma_0 + \int_{\Omega_0} \mathbf{f_0Z} d\Omega_0.
\]

Then by Gauss’s Theorem and the above equation we find that

\[
\mathbf{F_0Z} = \int_{\Omega_0} \left( \frac{\partial P_{ZX}}{\partial X} + \frac{\partial P_{ZY}}{\partial Y} + \frac{\partial P_{ZZ}}{\partial Z} \right) d\Omega_0 + \int_{\Omega_0} \mathbf{f_0Z} d\Omega_0.
\]

We can rewrite Reynold’s transport theorem (95) in the Lagrangian coordinate system, which is

\[
\frac{d}{dt} \int_{\Omega} \rho v_z d\Omega = \int_{\Omega} \rho \frac{DV_z}{Dt} d\Omega = \int_{\Omega_0} \rho_0 \frac{DV_z}{Dt} d\Omega_0.
\]

Note that \( \frac{DV_z}{Dt} = \frac{\partial V_z}{\partial t} \). Hence, we can rewrite Newton’s law in the Lagrangian coordinate system as

\[
\rho_0 \frac{\partial V_z}{\partial t} = \frac{\partial P_{ZX}}{\partial X} + \frac{\partial P_{ZY}}{\partial Y} + \frac{\partial P_{ZZ}}{\partial Z} + \mathbf{f_0Z},
\]

which is the equation of motion in the \( Z \)-direction. Then the equations of motion in the Lagrangian coordinate system are given by

\[
\rho_0 \frac{\partial v_X}{\partial t} = \frac{\partial P_{XX}}{\partial X} + \frac{\partial P_{XY}}{\partial Y} + \frac{\partial P_{XZ}}{\partial Z} + \mathbf{f_0X}, \\
\rho_0 \frac{\partial v_Y}{\partial t} = \frac{\partial P_{YX}}{\partial X} + \frac{\partial P_{YY}}{\partial Y} + \frac{\partial P_{YZ}}{\partial Z} + \mathbf{f_0Y}, \\
\rho_0 \frac{\partial v_Z}{\partial t} = \frac{\partial P_{ZX}}{\partial X} + \frac{\partial P_{ZY}}{\partial Y} + \frac{\partial P_{ZZ}}{\partial Z} + \mathbf{f_0Z},
\]

or written concisely as

\[
\rho_0 \frac{\partial \mathbf{V}}{\partial t} = \nabla \cdot \mathbf{P} + \mathbf{f_0}.
\]
Note that \( V = \frac{\partial \mathbf{U}}{\partial t} \). Hence, the Lagrangian equations of motion in terms of displacement is given by

\[
\rho_0 \frac{\partial^2 \mathbf{U}}{\partial t^2} = \nabla \cdot \mathbf{P} + \mathbf{f}_0.
\] (105)

**Remark 34.5** We note that the equations of motion (101) in the Eulerian (or moving) coordinate system are inherently nonlinear independent of the constitutive law assumptions (discussed in the next section) we might subsequently adopt. On the other hand, the Lagrangian formulation (105) (relative to a fixed referential coordinate system) will yield a linear system if a linear constitutive law is assumed. Thus, there are obvious advantages to using the Lagrangian formulation in linear elasticity theory (i.e., if a linear constitutive law is assumed).

### 34.3 Constitutive Relationships: Stress and Strain

Constitutive relationships or “laws” hold for a given material and are material dependent but independent of any coordinate system. However, we must always express these laws relative to the coordinate system in which we are working (e.g., see the Hooke’s law discussions below). In the preceding discussions, we have focused on relationships between displacements \( \mathbf{u} \) (or \( \mathbf{U} \)) (and their rates) and the stress \( \mathbf{S} \) (or \( \mathbf{P} \)). We have also related strain tensors \( \mathbf{E} \) to displacements. To complete our derivations of the equations of motion, we must assume (based on empirical observations) relationships (constitutive “laws”) between stress and strain.

For more details on our discussions in this section, one can refer to Fung’s Biomechanics ([18], pages 35-48).

**Definition 34.5** If a tensor has the same array of components when the frame of reference is rotated or reflected (i.e., invariance under rotation or reflection), then it is said to be an isotropic tensor.

A material whose constitutive equation is isotropic is said to be isotropic material.

**Remark 34.6** If the tensor \( D_{ijkl} \) is isotropic, then it can be expressed in terms of two independent constants \( \lambda \) and \( \mu \),

\[
D_{ijkl} = \lambda \delta_{ij} \delta_{kl} + (\delta_{ik} \delta_{jl} + \delta_{il} \delta_{jk}).
\] (106)

Now we shall discuss the three simple, idealized stress-strain relationships: nonviscous fluid, Newtonian viscous fluid, and Hookean elastic solid.
• Nonviscous fluid

\[ \sigma_{ij} = -p\delta_{ij}, \quad (107) \]

where \( \delta_{ij} \) is the Kronecker delta, i.e., \( \delta_{ij} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases} \), and \( p \) is a scalar called pressure.

**Note:** In an ideal gas, the pressure \( p \) is related to the density \( \rho \) and temperature \( T \) by the *equation of state*

\[ \frac{p}{\rho} = RT, \]

where \( R \) is the gas constant.

**Note:** For a real gas or fluid, it is often possible to obtain an equation of state

\[ f(p, \rho, T) = 0. \]

• Newtonian viscous fluid: a fluid for which the shear stress is linearly proportional to the strain rate.

\[ \sigma_{ij} = -p\delta_{ij} + D_{ijkl}V_{kl}, \quad (108) \]

where \( p \) is the static pressure (which depends on the density and temperature of the fluid according to the equation of state), \( D_{ijkl} \) is a tensor of viscosity coefficients of the fluid, and \( V_{kl} \) is the strain rate tensor.

For an isotropic constitutive relationship, we have

\[ \sigma_{ij} = -p\delta_{ij} + \lambda\delta_{ij}V_{kk} + 2\mu V_{ij}. \quad (109) \]

• Hookean elastic solid: a solid that obeys Hooke’s Law (in Eulerian coordinates)

\[ \sigma_{ij} = c_{ijkl}\varepsilon_{kl}. \quad (110) \]

If a material is isotropic, i.e., the array of elastic constants \( c_{ijkl} \) remains unchanged with respect to rotation and reflection of coordinates (\( c_{ijkl} = c_{jikl} \) and \( c_{ijkl} = c_{ijlk} \)), then the tensor \( c_{ijkl} \) can be written as

\[ c_{ijkl} = \lambda\delta_{ij}\delta_{kl} + \mu(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}). \quad (111) \]

Hence, by (110) we have

\[ \sigma_{ij} = \lambda\delta_{ij}\varepsilon_{kk} + 2\mu\varepsilon_{ij}, \quad (112) \]

where \( \lambda \) and \( \mu \) are *Lame’s parameters*, and \( \varepsilon_{ij} \) are defined by (82). Because Lagrangian coordinates are most often used in standard elasticity theory, it is useful to express Hooke’s law in Lagrangian coordinates. This is given by

\[ P_{ij} = C_{ijkl}\varepsilon_{kl}, \quad (113) \]
so that in the case of an isotropic material the stress-strain law in Lagrangian coordinates becomes

\[ P_{ij} = \hat{\lambda}\delta_{ij}\mathcal{E}_{kk} + 2\hat{\mu}\mathcal{E}_{ij}, \]  

(114)

with \( \mathcal{E}_{ij} \) given by (83), and \( \hat{\lambda}, \hat{\mu} \) are the Lame’s parameters in the Lagrangian coordinate system.
35 Information on Thumper Truck

The following information on a thumper truck is from the following website

- http://www.wisegeek.com/what-is-a-thumper-truck.htm

A thumper truck is a vehicle-mounted system for generating seismic vibrations and is used primarily by seismologists and geophysicists to locate and measure deposits of oil and natural gas. Thumper trucks are only one of several methods used to generate seismic data for this purpose, but they are the fastest, most efficient and least dangerous. Large and rugged, the thumper truck is designed for use in areas of difficult, often hilly terrain. It carries a heavy ground-impact weight which is raised by a hoist to a height of approximately nine to ten feet (2.7 to 3.04 m). The weight is then released to thump, or impact, the ground. Geophones, positioned to receive the seismic signal from the impact of the weight, then transmit the data to measuring instruments located in a recording vehicle nearby. This seismic information is often augmented by a series of impacts, either in the same spot or in several different locations nearby. This series of “thumps” is arrayed to enhance the quantity and accuracy of the data.

Though somewhat destructive in its own right, a thumper truck is far less environmentally devastating than dynamite, which, for some time, was the preferred method for the generation of seismic signal data. Dynamite is, by its nature, inherently dangerous to use. It is also ecologically devastating in that it necessitates the digging of “shot-holes” which are then expanded into sizeable craters by the explosive. This uncontrolled force can cause severe ecological damage over a much wider area than does a thumper truck. These vehicles maintain one further advantage over the use of dynamite in seismic data generation. They are a far less threatening exploration technique in areas of political instability than is any sort of explosive.

The thumper truck generates seismic data through a single impulse - the singular release of the weight. Vibroseis, a vehicle-mounted hydraulic vibrator, also known as a shaker unit, provides seismic signals over a longer time period, thus data is continuous as opposed to interrupted. Vibroseis vehicles are somewhat easier on the soil, as well. Specialized air guns and plasma sound sources (PSS), sometimes called spark gap sound sources, though less destructive than dynamite or a thumper truck, are not quite as accurate or as efficient.

Any exploration of natural deposits will be environmentally damaging to an extent. The use of a thumper truck is no different. USGS scientists estimate that the soil damage occasioned by the constant dropping of its impulse weight in a relatively constricted testing arc, especially along ridgelines, could take decades to rectify naturally.
36 1-D Problem

We discuss a careful formulation of a mathematical model for the problem of locating a buried metallic shell in an elastic medium (soil) using electromagnetic waves for interrogation. Impulsive forces are used to induce vibrations in the soil which will produce a vibrating target. Upon interrogation with the electromagnetic waves, a Doppler shift will be observed in the signal reflected from the vibrating shell.

For simplicity we assume initially that the target is buried directly beneath the impulse generator so that the elastic waves will travel vertically down the z-axis to impinge on and vibrate the target. In other words, we just focus on P-waves, propagating in the z-direction and generating shell displacements in the same direction. We take the location of the thumper at \( z = z_{p0} \) with the center of mass of target located at \( z = z_{10} \). The schematic of this problem is depicted in Figure 51.

![Figure 51: Buried metallic target problem in the field setting.](image)

37 Elastic Problem

For our initial studies, we assume that both soil and target are uniform in \( x \)- and \( y \)-directions, which implies that the contacting surface area between the target and the soil under the target is the same as the one between the target and the soil above the target. Note that the soil above the target is not as compact as that under the target. Hence, the material properties of soil above the target are different from those of soil under the target. For convenience, we assume that the soil above the target and below the target is homogeneous in the \( z \) direction (i.e., all the material properties of soil such as density of soil are assumed to be piecewise constant functions), and the target is homogeneous in the \( z \)-direction. In addition, we assume that soil behaves as a Kelvin-Voigt material for small vibrations: stress components in soil can be expressed as sum of two terms, the first term being proportional to the strain (\( \varepsilon \)) and the second terms being proportional...
to the rate of changes of strain ($\dot{\varepsilon}$). Since the target is metallic, we can treat it as a rigid body, which implies that the displacement of upper side of the target is exactly the same as that of its lower side. Hence, in the 1-dimensional setting we can treat the target as a point mass. With all the given assumptions, we can visualize the soil-target-soil as two thin rods connected by a point mass at $z = z_{10}$. The schematic of the problem is illustrated in Figure 52.

Figure 52: Schematic of 1-D setting for buried metallic target problem in the field setting.

Let $u(z, t)$ denote the displacement ($m$) in the $z$-direction at position $z$ at time $t$. Then we have

$$
\rho(z) \frac{\partial^2 u(z, t)}{\partial t^2} = \frac{\partial}{\partial z} \left( \kappa(z) \frac{\partial u(z, t)}{\partial z} + \eta(z) \frac{\partial^2 u(z, t)}{\partial t \partial z} \right), \quad z \in (z_{p0}, z_{10}) \cup (z_{10}, \infty),
$$

$$
M \frac{\partial^2 u(z_{10}, t)}{\partial t^2} = S \left[ \left( \kappa(z_{10}^+) \frac{\partial u(z_{10}, t)}{\partial z} + \eta(z_{10}^+) \frac{\partial^2 u(z_{10}, t)}{\partial t \partial z} \right) 
- \left( \kappa(z_{10}^-) \frac{\partial u(z_{10}, t)}{\partial z} + \eta(z_{10}^-) \frac{\partial^2 u(z_{10}, t)}{\partial t \partial z} \right) \right],
$$

(115)

Here $\rho$ denotes density of soil ($kg/m^3$), $\kappa$ is the elastic modulus of soil ($\frac{kg}{m^3 s^2} = Pa$), $\eta$ represents the damping coefficient of soil ($\frac{kg}{m^3 s}$), and they are all piecewise constant functions and given as follows:

$$
\rho(z) = \begin{cases} 
\rho^-, & z \in [z_{p0}, z_{10}) \\
\rho^+, & z \in (z_{10}, \infty)
\end{cases}, \quad \kappa(z) = \begin{cases} 
\kappa^-, & z \in [z_{p0}, z_{10}) \\
\kappa^+, & z \in (z_{10}, \infty)
\end{cases}, \quad \eta(z) = \begin{cases} 
\eta^-, & z \in [z_{p0}, z_{10}) \\
\eta^+, & z \in (z_{10}, \infty)
\end{cases}.
$$
For the second equation in (115), $M$ is used to denote the mass of target (kg) and $S$ represents the surface area of contact between the target and the soil under (or above) the target ($m^2$).

The **initial conditions** (assume zero displacement and zero velocity at $t = 0$) are given by

$$u(z, 0) = 0, \quad \frac{\partial u}{\partial t}(z, 0) = 0. \quad (116)$$

The **boundary condition** at $z = z_{p0}$ is given by

$$\left( \kappa(z) \frac{\partial u(z,t)}{\partial z} + \eta(z) \frac{\partial^2 u(z,t)}{\partial t \partial z} \right) \bigg|_{z=z_{p0}} = -f(t), \quad (117)$$

where $f$ is the applied external force with unit $N/m^2$.

### 38 Numerical Technique

To numerically solve (115)-(117) with finite elements methods (our choice here for mathematical and computational reasons), it is convenient to formulate the methods on a finite spatial domain. We do this by introducing an arbitrary lower (right) boundary at $z = z_{00}$. We do this in a manner so that it does not affect our solution at any time during the time period $t \in [0, T]$ of interest. For these calculations we care only about the displacement and velocity of the target in response to the input force and hence we will choose the right boundary $z_{00}$ to be sufficiently far beyond the target so that the moving elastic wave launched by the input force at $z_{p0}$ does not yet reach the boundary at $z_{00}$ in $t \in [0, T]$. Hence the displacement and velocity at $z_{00}$ will be zero in the time interval of our computations. This implies that we can set the boundary conditions

$$u(z_{00}, t) = 0. \quad (118)$$

Henceforth, our problem (115)-(117) is defined on the finite space domain $[z_{p0}, z_{00}]$. The numerical technique we use is the finite element method, which, as posed here only involves local elements so that no approximations are involved in setting the boundary conditions (118). To use the finite element method, it is most convenient to first rewrite (115) coupled with (117) and (118) in weak form.

#### 38.1 Weak Form

Let $< , >$ denote the usual $L^2(z_{p0}, z_{00})$ inner product, ie., $< \phi, \zeta > = \int_{z_{p0}}^{z_{00}} \phi(z) \zeta(z) dz$. Then by (115), (117) and (118), for any $\phi \in V = \{ \zeta | \zeta \in H^1(z_{p0}, z_{00}), \zeta(z_{00}) = 0 \}$ we
find that

$$
\left\langle \rho \frac{\partial^2 u}{\partial t^2}, \phi \right\rangle = \left\langle \frac{\partial}{\partial z} \left( \kappa \frac{\partial u}{\partial z} + \eta \frac{\partial^2 u}{\partial \partial z^2} \right), \phi \right\rangle
$$

$$
= \left. \left( \kappa(z) \frac{\partial u(z,t)}{\partial z} + \eta(z) \frac{\partial^2 u(z,t)}{\partial \partial z^2} \right) \phi \right|_{z_{10}^{-}}^{z_{10}^{+}} + \left. \left( \kappa(z) \frac{\partial u(z,t)}{\partial z} + \eta(z) \frac{\partial^2 u(z,t)}{\partial \partial z^2} \right) \phi \right|_{z_{10}^{-}}^{z_{00}^{+}}
$$

$$
- \left\langle \kappa \frac{\partial u}{\partial z} + \eta \frac{\partial^2 u}{\partial \partial z^2}, \frac{\partial \phi}{\partial z} \right\rangle
$$

$$
= -\phi(z_{10}) \frac{M}{S} \frac{\partial^2 u(z_{10}, t)}{\partial t^2} - \left\langle \kappa \frac{\partial u}{\partial z} + \eta \frac{\partial^2 u}{\partial \partial z^2}, \frac{\partial \phi}{\partial z} \right\rangle + f(t) \phi(z_{10}).
$$

Thus we have the weak form

$$
\left\langle \rho \frac{\partial^2 u}{\partial t^2}, \phi \right\rangle = -\phi(z_{10}) \frac{M}{S} \frac{\partial^2 u(z_{10}, t)}{\partial t^2} - \left\langle \kappa \frac{\partial u}{\partial z} + \eta \frac{\partial^2 u}{\partial \partial z^2}, \frac{\partial \phi}{\partial z} \right\rangle + f(t) \phi(z_{10})
$$

for any $\phi \in V$. 

128
MEASUREMENT OR OBSERVATION OPERATORS $f_j(q)$

- **Transport (mass, energy (heat), chemical species)**
  
  Mass $m(t), m(t, x)$; concentration or mass density $\rho(t), \rho(t, x)$; temperature $u(t, x)$, heat or temperature flux $\frac{\partial u}{\partial x}(t, x), \nabla u(t, x) \cdot \vec{n}$

- **Mechanical systems (force, momentum)**

  Displacement $y(t), u(t, x)$ with proximity probes; velocity $\frac{dy}{dt}(t), \frac{du}{dt}(t, x)$ with laser vibrometer; material strain $\frac{\partial u}{\partial x}(t, x)$ with strain guage; acceleration $\frac{d^2 y}{dt^2}(t), \frac{\partial^2 u}{\partial t^2}(t, x)$ with accelerometer; accumulated strain $\int \frac{\partial^2 u}{\partial x^2}(t, x) dx$ with piezoceramic patches

- **Populations (animals, molecules, cells)**

  Number $N(t), p(t), N(t, x), p(t, x)$; size density $u(t, x)$

**DATA:** $y_j, y_{jk}$ for $f(t_j; q), f(t_j, x_k; q)$
References


[123] M. Davidian, Marie’s ST 762 notes, Chapters 9 and 11.


