This work considers applied optimization in several contexts. Generally, we consider the optimal design of Traveling Wave Tubes (TWTs) and electron guns and the application of optimal control to human immunodeficiency virus (HIV) treatment schedules.

First, we consider the optimal design of electron guns using the 3D finite element code Beam Optics Analysis (BOA). Specifically, we consider the design of a Pierce gun that will be used in a 10MW multiple beam klystron at L-Band and a sheet beam gun (SBG). We consider various design considerations including beam shape, current, and electric field properties.

Secondly, we consider the design of TWTs, which are vacuum electron devices used for the amplification of radio frequency power. We use the simulation code CHRISTINE-1D to consider the design of a linear, C-Band, helix TWT and a folded waveguide slow-wave circuit.

Finally, we consider the problem of using optimal control methodology to design HIV treatment schedules. We first introduce a model describing HIV dynamics and study this model under optimal control based treatment strategies using a variety of sensitivity equations. We then study the Extended Kalman Filter (EKF) as a tool to filter data and estimate unknown states and model parameters. We then apply a Receding Horizon Control (RHC) methodology to the control of this model. This methodology is used in conjunction with the EKF, in the presence of noisy data, poor drug adherence and imperfectly known model parameters. We then study this methodology on a model of HIV dynamics that incorporates drug resistance.
Optimal Control, Estimation, and Shape Design: Analysis and Applications

by

John A. David

A dissertation submitted to the Graduate Faculty of North Carolina State University
in partial fulfillment of the requirements for the Degree of Doctor of Philosophy

APPLIED MATHEMATICS

Raleigh, North Carolina
2007

Approved By:

H.T. Tran
Chair of Advisory Committee

H.T. Banks

P. Gremaud

K. Ito

Z. Li
Dedication

This work is dedicated to my family, especially to the memory of my Dad.
Biography

John David was born on May 1, 1980 in Fort Worth, Texas to Joseph and Jeannie David. He has two younger siblings Julie and James. After a stint in Blacksburg VA, where his father earned a Ph.D in Mechanical Engineering, the family moved to Raleigh, NC. In 1998 he graduated from Millbrook High School. He then attended the University of North Carolina at Chapel Hill, graduating with Highest Distinction with a Bachelor of Science Degree in Mathematics and minor in English in May 2002. He then enrolled at North Carolina State University receiving a Master of Science Degree in Applied Mathematics in 2004. Upon the completion of this dissertation, he will have earned his Doctorate of Philosophy in Applied Mathematics. Beginning September, 2007, he will be employed by MIT Lincoln Laboratories in their Advanced Systems Concepts group.
Acknowledgements

I could not have finished this dissertation without an abundance of help. There’s no way I can adequately express the gratitude I owe to many people, but here goes nothing.

First, I would like to thank my advisor Dr. Hien Tran. I am grateful for his guidance, knowledge, patience and support. He is an amazing teacher and mentor, but a better person.

I also want to thank Dr. Thomas Banks, Dr. Pierre Gremaud, Dr. Kazufumi Ito and Dr. Zhilin Li for serving on my committee. Dr. Jesus Rodriguez, Dr. Stephen Campbell and Dr. Kris Jenssen deserve thanks for their contribution to my mathematical education.

I would like to thank Calabazas Creek Research, Inc. for their collaboration and support. I owe Dr. Lawrence Ives a great deal of thanks for letting me work on the problems involving electronic device design and offering the support and guidance to solve them. Thuc Bui, Dr. Carol Kory and Dr. Michael Read deserve thanks for their role in completing this research.

I am thankful to Dr. Paul Schlosser and CIIT Centers for Health Research for supporting me as a cooperative student and summer intern for two years and introducing me to research in applied mathematics.

I would be mistaken if I didn’t thank all the funding opportunities this work has given me. The electron gun design was funded by U.S. Department of Energy Small Business Innovative Research Grant DE-FG02-06ER86267. The TWT work was supported by the Naval Research Laboratory, contract number N00014-04-M-0170, and the Air Force Research Laboratory, contract number F29601-03-C-0049. The work related to the study of HIV was supported by grant NIH 1 R01 GM67299-04 and NIH/NIAID 9R01AI071915-05.

I couldn’t have done this without my family. I want to thank Mark Finley for being a great friend, reminding me to eat dinner and giving me somewhere to live for two years. Thank you James and Julie for being two of my best friends and always (sometimes) laughing at my bad jokes. Mom, you’ve always been my biggest fan and supporter and anything I accomplish is because of you. Jenny, thanks for putting up with and loving me as I finished this work, you truly bless my life and I love you very much.
Finally, deepest gratitude belongs "to him who is able to keep you from falling and to present you before his glorious presence without fault and with great joy to the only God our Savior be glory, majesty, power and authority, through Jesus Christ our Lord, before all ages, now and forevermore! Amen." Jude 24-25
# Contents

<table>
<thead>
<tr>
<th>List of Tables</th>
<th>vii</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Figures</td>
<td>vii</td>
</tr>
</tbody>
</table>

1 Introduction

2 Electron Beam Optimization
   2.1 Introduction        4
   2.2 Computer Tools      7
   2.3 Model Equations     10
   2.4 Design Iteration Procedure 11
   2.5 Optimization Routines  12
   2.6 Results            13
      2.6.1 Simple Parameters (cathode radius of curvature) 13
   2.7 Spline Cathode      18
   2.8 Electric Field Optimization: Spline Focus Electrode 22
   2.9 Sheet Beam Gun     26

3 TWT Optimization
   3.1 Introduction        32
   3.2 Design Of A Linear, C-Band, Helix TWT 35
   3.3 Designing Folded Waveguide Slow-Wave Circuits 42
   3.4 Optimal Design Results 45

4 HIV Model Analysis
   4.1 Introduction        48
   4.2 HIV Model           50
   4.3 Theoretical Considerations for Solutions to Model Equations 51
   4.4 Optimal Control: HIV Model  55
   4.5 Numerical Optimal Control  57
   4.6 Sensitivity Equations 61
      4.6.1 Sensitivities: HIV 64
List of Tables

2.1 Performance comparison for initial optimization, where the second column indicates the performance of the gun using the initial design and the third column indicates the performance of the gun with the optimized design parameters. ................................................. 17
2.2 Performance comparison for initial spline optimization. ...................................................... 22
2.3 Relative cost function values before and after optimization. .......................................... 29
2.4 Beam characteristics achieved versus goals. ..................................................................... 31
3.1 Nominal electron gun parameters. .................................................................................. 36
3.2 Helix parameters common to all optimizations. ............................................................... 36
3.3 Comparison of Nelder-Mead and steepest descent for random initial iterates. ............... 38
3.4 Results of DIRECT optimization for various combinations of weighting coefficients. ................. 46
3.5 Optimization results using the hybrid approach (DIRECT followed by Nelder-Mead). .......... 46
3.6 Frequency dependent range of saturated RF efficiencies, small signal gain variation, range of saturated output power and range of phase differences between small-signal and saturation. The non-optimized case, where the circuit period is uniform as a function of axial distance (no taper), and the optimized case, $a = b = 1, e = 5$, are compared. ............................................................. 47
4.1 Values of parameters in the HIV model. .......................................................... 51
4.2 $L^2$ ranking and norm of sensitivities of parameters. The first number in each column is the ranking of each parameter with respect to the data, with 1 being the most sensitive and 20 being the least sensitive. The number in the parentheses is the actual $L^2$ norm. Note that the second horizontal line separates the parameters in the equation for the immune response. ........................................................... 66
5.1 Parameters for RHC problem. ..................................................................................... 96
6.1 Values of parameters in the HIV model. ................................................................. 114
6.2 Parameters for RHC problem. ................................. 124
List of Figures

2.1 Pierce electron gun. .................................................. 7
2.2 Pierce electron gun in SolidWorks®. This is the 3D model of the gun shown in Figure 1. The anode is semi-transparent. ........................................ 8
2.3 Spherical cathode sketch in Solidworks® with associated design table. . 9
2.4 Flowchart for local optimization routine. ................................ 11
2.5 Sketch of cathode with associated design table. .......................... 14
2.6 Cross section of the cathode-anode region with associated design table controlling cathode to anode spacing. ......................................... 14
2.7 Examples of potential beam trajectories. ................................... 16
2.8 Design parameters as a function of iteration. .............................. 17
2.9 Goal functions as a function of iteration. ................................... 18
2.10 Optimization results. .................................................... 19
2.11 Spline cathode. .......................................................... 20
2.12 Sketch of cathode defined by spline curve. ............................... 21
2.13 Spline optimization: Before. ............................................. 22
2.14 Spline optimization: After. ................................................. 23
2.15 Initial focus electrode. Note that the dots along the focus electrode are the values specified by the user for the shape. The dark line is then the spline created from these points. The dashed lines specify other aspects of the CAD drawing. .................................................. 24
2.16 Optimized focus electrode. ................................................. 25
2.17 Final focus electrode. ...................................................... 27
2.18 Sheet beam gun in Solidworks®. ........................................... 28
2.19 Iteration history of SBG optimization. ..................................... 30
2.20 Particle trajectories in $xz$ and $yz$ planes, respectively. .................. 30

3.1 Schematic of a TWT. ..................................................... 33
3.2 Convergence of the helix pitch at fourteen axial grid locations from CHRISTINE-1D optimization as a function of iteration (each line represents pitch at a separate axial location). Convergence of the helix pitch at fourteen axial grid locations from the Nelder-Mead algorithm as a function of function evaluation (each line represents pitch at a separate axial location).

3.3 Comparison of CHRISTINE-1D’s native gradient-based optimization algorithm and the Nelder-Mead algorithm for maximizing the electronic efficiency.

3.4 Comparison of optimal helix pitches found by CHRISTINE-1D and Nelder-Mead.

3.5 Flowchart of folded waveguide optimization.

4.1 Optimal control: \( Q_e = 0.1, R_1 = 10, R_2 = 10, S = 1000, \epsilon_{1,max} = 0.7 \) and \( \epsilon_{2,max} = 0.3 \).

4.2 Model simulations under optimal control methodology.

4.3 Difference between scaled and unscaled sensitivity for \( d_1 \).

4.4 Relative sensitivities of parameters to total CD4 count, i.e., \( \frac{\partial T_1}{\partial q} T_1 + \frac{\partial T^*_1}{\partial q} q^* \).

4.5 Relative sensitivities of initial conditions to total CD4 count, i.e., \( \frac{\partial T_1}{\partial x_0} x_0 + \frac{\partial T^*_1}{\partial x_0} x_0^* \).

4.6 Generalized sensitivities to \( E \) under optimal control based treatment.

4.7 Generalized sensitivities to \( T_1 + T^*_1 \) under optimal control based treatment.

4.8 Filtering of noisy data, data taken daily. This filtering is the average over 20 runs. The plots on the right are the final 100 days of the filtering process.

4.9 Estimation of all states, data is taken daily. This filtering is the average over 20 runs. The plots on the right are of the final 100 days of the process.

4.10 Error between average estimate and true state, data taken daily. This estimation is the average over 20 runs. The plots on the left show the first 50 days of the process and the plots on the right show all of the simulation.

4.11 Filtering of noisy data, data taken daily. This filtering is the average over 20 runs. The plots on the right are the final 100 days of the filtering process.

4.12 Estimation of all states, data taken daily. This filtering is the average over 20 runs. The plots on the right are the final 100 days of the filtering process.

4.13 Error between average estimate and true state, data taken daily. This estimation is the average over 20 runs. The plots on the right are the final 100 days of the process.

4.14 Estimation of parameters, data taken daily. This estimation is the average over 20 runs. The plots on the left illustrate the first 50 days of this process. The plots on the right show the parameter estimation over the entire time course.
4.15 Filtering of noisy data, data taken daily. This estimation is the average over 20 runs. ................................................................. 87
4.16 Estimation of all states, data taken every 5 days. This estimation is the average over 20 runs. ................................................................. 88
4.17 Error of true states, data taken every 5 days. This estimation is the average over 20 runs. ................................................................. 89
4.18 Estimation of parameters, data taken every 5 days. This estimation is the average over 20 runs. ................................................................. 90

5.1 Schematic diagram of RHC methodology. ................................................................. 93
5.2 Example of RHC. ........................................................................................................ 95
5.3 Model response and control when the control is allowed to vary daily. ................ 98
5.4 Phase plot of log of immune effectors vs. log of virus when control is allowed to vary daily on the left. The results for when control was allowed to vary every 10 days are on the right. ................................................................. 99
5.5 Model response and control when the control is allowed to vary every 5 days. ........................................................................................................ 99
5.6 Model response and control when the control is allowed to vary every 10 days. ........................................................................................................ 100
5.7 Model response and control when control is allowed to vary daily. Initial conditions given by unhealthy steady state. ......................................................... 100
5.8 Phase plot of log of immune effectors vs. log of virus when control is allowed to vary daily and the initial condition is the unhealthy equilibrium. ................................................................. 101
5.9 Model response and control when control is allowed to vary every 5 days. Initial conditions given by unhealthy steady state. ......................................................... 101
5.10 Model response and control when control is allowed to vary every 10 days. Initial conditions given by unhealthy steady state. ......................................................... 102
5.11 Model response and control when control is allowed to vary daily and control was interrupted after 300 days for 10 days. ................................................................. 103
5.12 Immune response when control is allowed to vary daily. Comparison between following treatment and when treatment was interrupted after 300 days for 10 days. Plot on the left contains the full time interval, while the plot on the right focuses on the time period directly after the patient went off medication. ................................................................. 103
5.13 Immune response when control is allowed to vary every 5 days. Comparison between following treatment and when treatment was interrupted after 300 days for 10 days. Plot on the left contains the full time interval, while the plot on the right focuses on the time period directly after the patient went off medication. ................................................................. 104
5.14 Control found when EKF is used as state estimator. Performance of EKF as a filter of noisy data. ................................................................. 105
5.15 Performance of EKF as a state estimator. Red line denotes EKF estimation, while blue line represents true state. .................. 106
5.16 Control found when EKF is used as state estimator. Performance of EKF as a filter of noisy data. Unscheduled 10 day treatment lapse after 300 days. ............................................. 107
5.17 Performance of EKF as a state estimator. Red line denotes EKF estimation, while blue line represents true state. Data was collected every 5 days. Unscheduled 10 day treatment lapse after 300 days. ............ 108
5.18 Model response and control when control is allowed to vary daily and model parameters are randomly perturbed. ................... 110

6.1 Only PIs administered. ...................................................... 119
6.2 Only RTIs administered. .................................................... 120
6.3 Model response for drug-resistant model when control is allowed to vary daily. .......................................................... 125
6.4 Control for drug-resistant model when the control is allowed to vary daily. ................................................................. 126
6.5 Model response for drug-resistant model when control is allowed to vary every 5 days. .................................................... 127
6.6 Control for drug-resistant model when control is allowed to vary every 5 days. ............................................................ 128
6.7 Model response for drug-resistant model when control is allowed to vary daily and control was interrupted after 200 days for 10 days. ......... 129
6.8 Control for drug-resistant model when control is allowed to vary daily and control was interrupted after 200 days for 10 days. ................. 130
6.9 Comparison of immune effector response when control was interrupted and when it was followed. ................................. 131

7.1 Example of potential design for MBG. ............................... 136

A.1 Nelder-Mead in 2 dimensions. ............................. 151
A.2 Example of 1D objective function in which implicit filtering would be useful. .................................................... 152
A.3 An illustrative example of the iterative process of the DIRECT algorithm in a plane. The shaded rectangle is the region that is identified by the algorithm for division. .......................... 154
Chapter 1

Introduction

When one is attempting to cause a system, be it physical, economic, biological, etc. to behave in some manner, the question that often arises is: what is the best way in which this system can operate and how must we manipulate the system in order for it to operate in this manner? When formulated mathematically we consider this question to be an optimization problem. The process to posing and solving this problem generally involves several steps. First, one must have a mathematical description of how the system behaves. This step is generally considered model development and is of itself a complex and challenging field. There are a variety of classes of models including continuous, discrete, deterministic, and stochastic. In this work we will consider deterministic models governed by ordinary and partial differential equations. One must also be able to mathematically describe what sort of performance is desired. This step is generally described by some sort of goal or cost function. One must also understand how to solve at least approximately the model equations. Since optimization problems generally require multiple simulations of these equations the solutions of these simulations must be able to be acquired relatively quickly. As computational efficiency has increased so has the ability to consider more complex models for these physical/biological systems. Finally, one must consider what sort of optimization routines are appropriate for the given problem.

This thesis will study various aspects of this process in the context of a few specific applications. First, we will consider the design of 3D electronic devices specifically electron beam devices. These devices are increasingly being designed in a manner that is
inherently 3D, i.e., they cannot be accurately modelled with lower dimensional approximations. Not only does this require the development of accurate simulation codes but it increases the demand on design engineers. As the design geometries become more complex and the numerical simulations become more costly we must develop ways to simplify and improve on the design procedure. We will address this with the use of numerical optimization.

Next, we will begin by considering the problem of designing Traveling Wave Tube amplifiers. TWTs are vacuum devices invented in the early 1940’s for amplification of radio frequency (RF) power. These devices are critical for radar, communications and electronic warfare missions in the military, as well as in commercial applications. The physics-based design and simulation code, CHRISTINE-1D, was used in the past to explore different TWT circuit designs and to automate the process of parameter estimation. However, the current capability of CHRISTINE-1D allows optimization of only helix TWT designs, and includes a limited number of optimization goal functions. In addition, the current optimizer in CHRISTINE-1D employs a modified steepest descent method to carry out the optimization process. We will attempt to improve upon the results in the design of these devices using various optimization routines and goal functions.

Finally, we will use optimization to consider the treatment of individuals infected with HIV. According to the World Health Organization there are an estimated 39.5 million people infected with HIV and there were 4.3 million new infections in 2006 (WHO 2006). Our research considers the integration of modelling, optimization, and estimation to design treatment schedules that can control the infection. Specifically, we will consider the treatment problem as an optimal control problem, and then progress to considering feedback control through the use of RHC. As feedback requires state estimation we will consider the use of stochastic estimation, specifically the EKF. As a response to the growing concern about drug-resistant strains of this virus we will consider an alternate model and apply our control methodology.

The remainder of the dissertation is arranged as follows. Chapter 2 considers the problem of designing electron guns using computer optimization techniques. Several different design parameters are manipulated while considering multiple design criteria including beam and gun properties. The equations modelling the electron beam dynamics
and the optimization routines are described. Examples of guns designed using these techniques are presented.

Chapter 3 considers the use of optimization in the design of TWTs. The objectives of our work are three-fold: (i) to investigate optimization techniques that may be better suited for this problem (for example, simplex type methods such as Nelder-Mead and DIRECT), (ii) allow optimization of non-helix TWTs, (iii) and implement new optimization goal functions. Finally, to show the feasibility of our approach, we apply our optimization algorithms to the problem of designing a folded waveguide slow-wave circuit.

Chapter 4 introduces a model of the immunologic response of HIV in individuals. It then illustrates how optimal control methodology can produce a drug dosing strategy and how this treatment strategy has features of Structured Treatment Interruption (STI). We then perform a sensitivity analysis of the model including both classical sensitivity functions and generalized sensitivity functions. We also study how stochastic estimation can be used to filter and estimate states and parameters from noisy data. In the course of this analysis it will be shown how Automatic Differentiation (AD) can be a powerful tool for this type of analysis.

Chapter 5 illustrates how a RHC methodology can be used to drive the system to a stable equilibrium in which a strong immune response controls the viral load in the absence of drug treatment. We also illustrate how this feedback methodology can overcome unplanned treatment interruptions. We consider how ideas from stochastic estimation can be used in conjunction with RHC to create a robust treatment methodology.

Chapter 6 then, considers an alternate model of HIV infection. This chapter describes control attempts in the case when drug resistant strains of the virus are considered.

Finally, in Chapter 7 we present a summary and future directions for this work.
Chapter 2

Electron Beam Optimization

2.1 Introduction

Electron guns are used in many vacuum electron devices to convert electrical power into an electron beam. Electron beam devices include RF sources for numerous applications such as communications, radar, industrial heating, and high energy accelerators. Electron beams are also used in medical and industrial x-ray devices, for electron beam lithograph and electron beam welding, and in cathode ray guns for televisions and oscilloscopes. Many of these devices are critical for national defense and science as well as industrial applications.

In an RF source, the beam energy is converted to energy in an RF wave. In x-ray sources, welders, or cathode ray devices, precise focusing is required to obtain high resolution. The electron gun is a primary component in klystrons, TWTs, gyrotrons, and inductive output tubes. The configuration of the gun depends on many factors, including the operating voltage, current, beam size and shape, magnetic focusing circuit, power supply, and operational environment. Consequently, customized electron gun design is required for essentially every new device. Because of the large number of variables, this is often a time consuming and expensive process. Typical design time for a new electron gun for an RF source is 30-40 man-hours involving 20-30 design iterations (Cattelino 2006). This is for designs that can be modelled in two dimensions. A number of 2D simulation tools are available, including EGUN® and TRAK®.
The process becomes more demanding for 3D designs. Recent interest in multiple beam and sheet beam guns is placing severe demands on the computational codes as well as the design engineer. 3D analysis is computationally intensive, making iterative design very expensive when performed manually. As an example, Calabazas Creek Research, Inc. (CCR) recently designed a sheet beam gun for an X-Band klystron (Read, Jabotinski, Miram & Ives 2005). The development required approximately 2-3 man-months with approximately 100 iterations, each involving 8 hours of CPU time. The gun operated with Brillouin focusing, so the simulations did not include a magnetic field.

CCR completed the design of an X-Band electron gun using a combination of MAFIA® and another 3D simulation code developed in Russia (Ives, Miram & Krasnykh 2004). This design was performed several years ago and required approximately one year of iterative design. Fortunately, more recent computational tools and experience are reducing this design time, though a recent design of a multiple beam gun for a 200 MHz Multiple Beam Klystron (MBK) still required four months of iterative design (Development of a 200 MHz Multiple Beam Klystron 2005). Both these efforts were for singly convergent guns. A recent program to design a doubly convergent multiple beam gun was abandoned because the number of variables and the complexity of the design could not be overcome. The general difference between these types of electron guns and the issues related to designing them will be discussed further in Chapter 7. These earlier efforts provided much of the motivation for developing computer optimized design.

In a typical design process, the engineer begins with an existing configuration close to the new requirement or with design equations for the basic configuration. The design is simulated and geometrical, electrical, and magnetic parameters are modified until the required performance is achieved. In general, the engineer attempts to design an electron gun operating with a specified combination of voltage and current that produces an electron beam of a specific size with laminar electron trajectories. For confined flow guns where the magnetic field penetrates to the cathode, the engineer must design both the electrical and magnetic circuits. The requirement is to emit the electrons at a precise angle to the magnetic field to minimize beam scallop (Gilmour 1994). Scallop is a measure of the difference between the maximum diameter of the beam and the minimum diameter of the beam. A typical requirement is that the beam envelope not vary more than 5%
through the device.

Previously, CCR developed a computer optimization process for Brillouin focused electron guns (Lewis, Tran, Read & Ives 2004). This research used EGUN® and MATLAB® routines to produce an electron beam of a specific size at a specified axial location with laminar electron trajectories. This work successfully demonstrated that significant reduction in the design time could be achieved with minimal interaction by the design engineer. This translates into a reduction in design cost and potentially improved performance.

In our current research we are advancing this capability using the 3D finite element, adaptive mesh code Beam Optics Analysis (BOA). Although the initial designs presented here are two dimensional, the simulations are completely three dimensional. While this is not necessary for these particular designs, it establishes the process for direct transition to the 3D designs described later. In this research, a confined flow, Pierce electron gun is designed using computer optimization. There were several goals in the development. The initial effort involves modification of simple geometrical parameters to achieve the optimum beam quality. These parameters included the spacing between the cathode and anode to achieve the desired perveance, the relative position of the magnetic circuit to achieve the desired beam compression, and the radius of curvature of the cathode to minimize beam ripple. Figure [2.1] shows a drawing of the electron gun and Figure [2.2] shows the same gun in SolidWorks®.

The next task is to demonstrate that nonlinear shaping of electrode surfaces could improve performance. This research focuses on two sub tasks. The first was to demonstrate that the shape of the cathode could be optimized to reduce beam ripple while still achieving the desired perveance and beam size. The cathode is defined by a set of points connected with a spline curve. This curve is rotated about the gun axis to define the cathode surface. The second task is to define the focus electrode by a similar set of points and a spline curve, then optimize the shape of the curve to minimize the electric field. Minimization of the electric field will reduce the probability of arcing between the gun and the anode. The final step in the design of the Pierce electron gun is to combine these results to demonstrate a high quality electron beam with minimal potential for electrical breakdown.

As a first test of the fully 3D electron gun design we will consider the design of a sheet
beam gun (SBG). Sheet beam cathodes typically use a cylindrical emitter to produce the rectangular electron beam.

The organization of the chapter is as follows. The gun geometry was generated in SolidWorks® with BOA used for the computer simulations. These tools are described in Section 2.2. The model equations are described in Section 2.3. Section 2.4 describes the process by which MATLAB® routines control the optimization process by executing both SolidWorks® and BOA in batch mode using line commands. Section 2.5 describes the optimization routines. Sections 2.6, 2.7 and 2.8 describe the simulation results. Section 2.9 details the initial design of a SBG.

### 2.2 Computer Tools

SolidWorks® and BOA are the principal commercial programs used in this research. SolidWorks® is a 3D, parametric, solid modelling program that can generate ACIS-formatted geometry files, a requirement for integration with BOA. Figure 2.2 shows the
Pierce electron gun in SolidWorks®. In this figure, the anode/beam tunnel is semi-transparent.

Figure 2.2: Pierce electron gun in SolidWorks®. This is the 3D model of the gun shown in Figure 1. The anode is semi-transparent.

An important feature of parametric modelling is the ability to define key dimensions in design tables. One can then update the geometry by changing values in these tables. This allows an external program to control parametric changes to the geometry. Figure 2.3 shows a sketch of a spherical cathode in SolidWorks® with the associated design table in Excel®. The cathode is created by revolving the sketch about the axis.

SolidWorks® allows batch operation, so the MATLAB® control program can modify the design table, execute the CAD program to generate the updated model, generate the ACIS file, then terminate the CAD program.

The electron gun simulation is performed using BOA. Like the solid modelling program, BOA can be executed by MATLAB®. All input for BOA is contained in an ASCII file that can be modified by the MATLAB® control program. The file contains information for solution of the electric fields, including the voltages assigned to various
objects and dielectric constants for ceramics. Electron emitters are also defined using information from the CAD program to identify specific surfaces. The ASCII file provides information controlling the number of trajectories and the temperature and work function for thermionic emitters. All meshing is performed automatically within BOA. For this research, the magnetic field profile was generated by Maxwell 2D® and used as input to BOA. The axial position of the magnetic circuit relative to the center of the cathode was a variable in the optimization process and controlled the beam compression. Magnetic circuit modelling will soon be implemented within BOA, so input from external programs will not be necessary. This will also allow optimization of the magnetic circuit parameters in future research.

BOA is an adaptive mesh, finite element, 3D analysis code for designing electron beam devices. A principal feature is the adaptive meshing which removes the burden for mesh generation from the user and assigns responsibility to the field solver and particle pusher routines. With adaptivity enabled by the user, BOA adapts the mesh density in areas where field gradients are high until the specified accuracy is achieved. It can also coarsen mesh in areas where high accuracy is not required to reduce the computational
burden. The user can also control the mesh density in regions occupied by the electron beam and in regions near selected surfaces.

2.3 Model Equations

The model equations for the electron guns are based on the time-independent Maxwell’s equations. The general process of computing the performance of the electron gun consists of solving for the electric field using Poisson’s equations and determining the particle trajectories using the Lorentz force equations.

First, to determine the electric field we solve

\[ \nabla^2 \Phi = -\frac{\rho}{\epsilon_0}, \]  

(2.1)

where \( \Phi \) is the electric potential, \( \rho \) is the charge density, and \( \epsilon_0 \) is the permittivity of vacuum. Note that once the electric potential is found the electric field, \( E \), is given by

\[ E = -\nabla \Phi. \]  

(2.2)

Once this is solved the electron trajectories are found using the Lorentz force equations give by

\[ \frac{dP}{dt} = q(E + v \times B), \]  

(2.3)

where \( P \) is the momentum, \( v \) is the velocity and \( B \) is the magnetic field. Note this calculation includes a relativistic correction so

\[ P = m\gamma v, \]  

(2.4)

where

\[ \gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} \]  

(2.5)

Overall, BOA iterates in the following manner:

1. The electric field is calculated without the space charge.

2. The trajectories of the particles under the influence of the electric field are calculated.
3. The electric field is recalculated including the effects of the space charge.

4. This is repeated until the change in current density at the emitter and the maximum potential field remain relatively unchanged between iteration.

### 2.4 Design Iteration Procedure

Each iteration of the optimization routine requires several steps. The general process is described in Figure 2.4; however, we will describe each block in more detail here.

![Flowchart for local optimization routine](image)

**Figure 2.4:** Flowchart for local optimization routine.

As usual, the iterative methods used in this research require a starting point or initial design. For each optimization attempt, the user must specify a set of starting design parameters for the optimization routines. It is generally beneficial if these design parameters are relatively close to the optimal design parameters; however this may not
be necessary. There are routines that can consider a general subset of the parameter space and attempt to find a global minimum, but these routines generally require an extensive number of function evaluations, which is not feasible in the case of 3D design.

The first step in a function evaluation is to write the geometry related parameters, e.g., cathode radius or spline parameters, focus electrode shape parameters etc., into Excel® files linked to the SolidWorks® CAD files. We used a routine written by Brett Shoelson, which was obtained at MATLAB® central, an open exchange for MATLAB® users, to write the numerical values from MATLAB® to the Excel® files. SolidWorks® then regenerates the geometry files with the newly updated parameters from the spreadsheets. This produces a geometry file read by BOA, which then executes, producing output files detailing the trajectories of the particles and fields in the electron gun. MATLAB® routines read these files to determine the beam characteristics and calculate a cost function value that measures how well the current design parameters achieve the design goals. Finally, the optimization routine uses this cost function value to either compute a new set of trial design parameters or, in the case that the current design parameters are considered optimal, to terminate the routine.

2.5 Optimization Routines

This section provides an overview of the sampling optimization algorithms used in the optimal design of the electron guns. Basically, in an optimal design problem, one begins by formulating a function that characterizes the design goals. The task is then to minimize or maximize this function and thus obtain a design that meets the desired criteria. Mathematically speaking, the problem is given a function $f : \mathbb{R}^N \rightarrow \mathbb{R}$ find $\lambda^* \in \mathbb{R}^N$ such that $f(\lambda^*) \leq f(\lambda)$ for all $\lambda$ of interest. If the $\lambda$’s of interest are only those near $\lambda^*$, then it is a local optimization problem. On the other hand, if the $\lambda$’s of interest belong to a subset $\Omega \subset \mathbb{R}^N$, then it is a global optimization problem. The Nelder-Mead and implicit filtering optimization routines used in this research are known as deterministic sampling methods. Gradient information used by implicit filtering is only approximate, as it is obtained from sampled points in the parameter space. For a further description of these methods see Appendix A. For a discussion on the advantages and
disadvantages of sampling based methods versus gradient based methods, the interested reader is referred to (Kelley 1999).

2.6 Results

The electron gun chosen for this research is a Pierce gun that will be used in a 10 MW, multiple beam klystron at L-Band. The specifications require that the beam voltage be less than 120 kV. Given the total power required and the number of guns anticipated, each gun should produce approximately 20 A. Based on a preliminary RF circuit design, a beam fill factor of 66% was chosen as the beam size, i.e., the beam radius should be 66% of the radius of the beam tunnel. Consequently, the goal is to develop an electron gun operating at 110 kV, producing 20 A with a 66% fill factor in the beam tunnel. An additional requirement is that the beam ripple be less than 5%.

2.6.1 Simple Parameters (cathode radius of curvature)

The first task was to optimize the gun performance by changing simple parameters in the model. For the cathode, the only variable was the radius of curvature. This is illustrated in Figure 2.5 showing the SolidWorks® sketch and the associated design table. In the optimization process, the radius of curvature was modified to match the magnetic field profile at the cathode to minimize beam ripple.

The gun current was modified by adjusting the spacing between the cathode and the anode. This is illustrated in Figure 2.6, which shows a cross section of the cathode-anode region and the associated design table. Since the face of the cathode is curved, the spacing indicated is from the back of the cathode to the flat face of the anode.

The third variable was the relative position between the cathode and the magnetic circuit. Since BOA’s magnetic field solver is not yet operational, the magnetic circuit was modelled in Maxwell 2D and the solution used as input to BOA. This input is part of the ASCII input file into the BOA solver, so the MATLAB® routines directly edit this file to adjust this parameter.

Evaluation of simulation results was performed in MATLAB® using output data
Figure 2.5: Sketch of cathode with associated design table.

Figure 2.6: Cross section of the cathode-anode region with associated design table controlling cathode to anode spacing.
generated by BOA. The beam current is reported by BOA directly. The beam size was evaluated from the radial coordinate of the outermost trajectory. Beam laminarity was determined as follows. Consider the radius, $r$, of each electron particle as a function of its $z$ location. Thus for each particle (disregarding the angular component, which does not affect the beam laminarity) one can characterize the $ith$ particle by the function $r_i(z)$. Then, a measure of the laminarity is the derivatives of this function with respect to $z$. However, it may misrepresent the true beam shape to only look at these values at a single $z$ location, so this information is determined at several locations along the length of the beam. Since we are looking at trajectory information at specific $z$ values, we have included information about first and second derivatives of the trajectories. A zero value of the first derivative is achieved at a peak or trough in the beam shape. However, the beam will be laminar at this point if the second derivative is also zero. See Figure 2.7 for the potential trajectories that caused us to use first and second derivative information.

The goal functions that describes beam laminarity are given by

$$J_1(\lambda) = \alpha \sum_{j=1}^{4} \sum_{i=1}^{N} \frac{\partial r_i(z_j)}{\partial z}$$

$$J_2(\lambda) = \beta \sum_{j=1}^{4} \sum_{i=1}^{N} \frac{\partial^2 r_i(z_j)}{\partial z^2}$$

where $\alpha$ and $\beta$ are user specified weights, and $N$ is the number of particles used. Note the upper bound of 4 in the sum over the $j$ index is the number of $z$-locations at which this information was obtained.

The second goal is to produce a beam with a specific radius. The goal function used here is

$$J_3(\lambda) = \gamma (b - b_d)^2$$

where $b$ is the beam radius, $b_d$ is the desired beam radius, and $\gamma$ is a user specified weighting parameter.

Finally, the beam should have a specified current. The appropriate goal function is

$$J_4(\lambda) = \delta (I - I_d)^2$$

where $I$ is the beam current, $I_d$ is the desired beam current, and $\delta$ is a user specified
weighting parameter. The total goal function is

\[ J(\lambda) = \sum_{i=1}^{4} J_i, \]  

(2.10)

where \( \lambda \) are the parameters that specify gun design.

Figure 2.7: Examples of potential beam trajectories.

The specific numbers for this optimization are as follows. The weights \( \alpha, \beta, \gamma, \) and \( \delta \) were chosen such that \( J_i(\lambda_0) = 1 \) for \( i = 1, .., 3 \) and \( J_4(\lambda_0) = 10 \), where \( \lambda_0 \) is the initial value for the design parameters. The initial value for this problem is \( \lambda_0 = (50.8 mm, 25.4 mm, 0 mm) \), where the entries in \( \lambda_0 \) are the spacing between the cathode and the anode, the radius of curvature of the cathode, and the relative position between the cathode and the magnetic circuit, respectively. The desired beam radius is \( b_d = 6.2 \) mm. The desired current \( I_d = 20 \) A. The optimization process used the Nelder-Mead algorithm. A 2.8 GHz Toshiba laptop running Windows XP executed the optimization routines, the CAD, and simulation codes. Each iteration in the process
required approximately eight minutes. Figure 2.8 shows the value of various design parameters during the process. Figure 2.9 shows the value of the various goal functions as the iterations proceed. The results of this optimization are summarized in Table 2.1.

![Graphs showing various parameters](image)

**Figure 2.8:** Design parameters as a function of iteration.

**Table 2.1:** Performance comparison for initial optimization, where the second column indicates the performance of the gun using the initial design and the third column indicates the performance of the gun with the optimized design parameters.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>λ₀ (initial values)</th>
<th>λ (optimized parameter)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beam Radius</td>
<td>4.855 mm</td>
<td>6.523 mm</td>
</tr>
<tr>
<td>$\frac{1}{3}J_2$</td>
<td>2.068e+022</td>
<td>5.354e+021</td>
</tr>
<tr>
<td>$\frac{1}{7}J_3$</td>
<td>4.662</td>
<td>1.311</td>
</tr>
<tr>
<td>Current</td>
<td>27.859 A</td>
<td>19.587 A</td>
</tr>
</tbody>
</table>

As can be seen, the variables appear to reach essentially their final values after approximately 60 iterations. This required approximately eight hours.

The results are summarized in Figure 2.10. In addition to the previously defined goal functions, the beam ripple or scallop is a value of interest. The beam scallop is defined
Figure 2.9: Goal functions as a function of iteration.

\[ Scall(\lambda) = 100 \frac{d_{\text{max}} - d_{\text{min}}}{d_{\text{max}} + d_{\text{min}}}, \]  

(2.11)

where \( d_{\text{max}} \) and \( d_{\text{min}} \) denote the maximum and minimum diameter of the beam. The current was within 2\% of the goal and the beam scallop was well within the goal of 5\% or less. The beam size was approximately 7\% from the desired value. Considering that the magnetic profile is not optimized and only three variables were available for modification, these results were very encouraging.

### 2.7 Spline Cathode

It is anticipated that optimization of 3D structures will require modification of electrode shapes that can not be achieved with simple dimensional changes of lines, arcs or circles. Consultation with CCR engineers indicated that outstanding results could be achieved if one is not restricted to simple changes of lengths and radii (Ives, Neilson & Vogler 1998, Neilson 2006). The next task was to demonstrate that shape optimization could be implemented and achieve further improvements in performance.
A cathode was designed that consisted of points along a radius of the cathode connected by a spline. The full cathode was generated by rotating the curve about the cathode axis. Figure 2.11 shows the initial configuration. The cathode was defined by six points along a constant curvature arc from the optimization in the previous section. The optimization routines were allowed to freely change the axial position of the points, but not the radial position.

For this optimization, the beam scallop was again evaluated as given by formula (2.11). The goal was to minimize the function

\[ J(\lambda) = J_3(\lambda) + Scall(\lambda). \] (2.12)

The initial decision to vary all points of the spline proved problematic. Changes to points near the axis could not impact the value of the cost function. Thus, the process failed to converge and produced unrealistic cathode shapes. Therefore, it was necessary to limit the number of points that could be varied, specifically, those near the outer edge, and constrain the range of modifications for each point. Consequently, the design
engineer can not be totally eliminated from decisions necessary to achieve a complete design.

Only the outer two points were selected for modification, and, again, only the axial coordinate could be changed. As a constraint, the axial distance of each point from the center of the cathode could not be less than any points with a smaller radius. This would result in electrons being emitted with a positive radial velocity, which was assumed to be at variance with the desired result. The distance between the cathode and anode was fixed from the previous optimization. The relative position between the cathode and the magnetic circuit was the third parameter in addition to the two outer points on the spline cathode.

For this problem the implementation Nelder-Mead did not vary the initial parameters much, so we switched to use the optimization routine implicit-filtering. The optimization resulted in a beam scallop of less than 5%, and the beam size was within 2% of the target value. The final cathode shape is shown in Figure 2.12. Note the subtle deviation from a purely spherical shape near the outer edge. The results are summarized in Table 2.2.
Figure 2.12: Sketch of cathode defined by spline curve.
Table 2.2: Performance comparison for initial spline optimization.

<table>
<thead>
<tr>
<th>BeamRadius</th>
<th>$\lambda_0$ (initial values)</th>
<th>$\lambda$ (optimized parameter)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale</td>
<td>4.25mm</td>
<td>6.21 mm</td>
</tr>
<tr>
<td>BeamRadius</td>
<td>16 %</td>
<td>4.83 %</td>
</tr>
</tbody>
</table>

and Figures 2.13 and 2.14.

2.8 Electric Field Optimization: Spline Focus Electrode

A key failure mechanism in electron guns is arcing between the focus electrode and the anode. Consequently, there is a desire to minimize the electric field so long as the beam performance is not compromised. This chiefly involves keeping the appropriate angle between the cathode and the focus electrode while shaping the focus electrode. For this optimization, the focus electrode was defined by a series of points connected
with a spline. As with the cathode, the final electrode was generated by rotating the curve about the axis. Figure 2.14 shows the starting configuration for the optimization which consisted of a straight line between the inner radius of the electrode to a constant radius arc tangent to the outer radius. In discussion with CCR engineers we learned that high field gradients only occur in the region of the focus electrode closest to the anode. Consequently, only points in this region, specifically, points P4 through P6, were selected for optimization. The optimization routines were allowed to change the axial position of the points, but not the radial position. Constraints were added to prevent generation of unrealistic configurations.

BOA can calculate electric field vectors at discrete points on the surfaces. The electric field vector on the focus electrode was denoted by $E_i$. The goal function for this problem was

$$J(\lambda) = \max ||E_i||.$$  \hspace{1cm} (2.13)

Note that since the goal function is the maximum value of the electric field on the focus electrode and we are trying to minimize this goal function, we are thus minimizing the
Figure 2.15: Initial focus electrode. Note that the dots along the focus electrode are the values specified by the user for the shape. The dark line is then the spline created from these points. The dashed lines specify other aspects of the CAD drawing.
maximum value of the electric field in the device. Using the Nelder-Mead algorithm, the magnitude of the vector was reduced for $J$ by 6.7%. Figure 2.16 shows the final shape from the optimization process.

![Optimized focus electrode](image)

Figure 2.16: Optimized focus electrode.

Recall that only points P4 through P6 were modified during the optimization process. The position of P3 was not allowed to change, which explains the curvature in the profile at this location. While this shape satisfies completely the optimization goal, it could present unnecessary complications for machining. Consultation with CCR engineers taught us that slight changes in this region would not significantly impact perfor-
mance, therefore the location of P3 was manually modified to remove the slight bump in the profile. This is another indication that engineering experience and knowledge are important in any computational design effort. The resultant change in the maximum electric field was less than .008%. This final configuration is shown in Figure 2.17. The final task was to ensure that the focus electrode modification was still compatible with the desired beam performance. The full beam simulation was performed, and the change in beam diameter was negligible, and the beam scallop changed from 4.8% to 5.1%. As this violated our goal of under 5% scallop, we reran an implicit filtering optimization similar to that described in Section 2.7 with the focus electrode found in this section. Using this routine we were able to reduce the beam scallop to 4.4068%.

2.9 Sheet Beam Gun

As a first test of a true 3D electron gun design we will attempt to design a SBG. Several organizations, including CCR, are developing sheet beam RF sources. These have lower operating voltage, improved efficiency, greater bandwidth, and reduced fabrication cost. In high power RF sources, high power beams are required. SBGs can distribute the beam over a larger area than a round electron beam. This reduces the space charge in the beam. Since the beam has a smaller space charge, one can produce a beam with a smaller voltage. If the current of the beam is then increased the beam will produce the same power. The beam with reduced voltage then requires a smaller voltage source as well as less shielding for the x-ray radiation the higher voltage produces. They are being developed for the International Linear Collider, which will require several hundred klystrons (Ives 2007). Therefore, the potential impact is quite large. The specific gun we will design will operate at 50kV and a picture of the geometry is depicted in Figure 2.18.

The parameters that we can manipulate in the design of this gun are the cathode to anode spacing, the cathode radius and the angle of the focus electrode. After discussion with CCR engineers we have decided that the desired gun performance would be one in which the current is 11A, the beam does not compress in the $xz$ plane (thus having a width of 100mm) of the device and we want the minimum beam width in the $yz$ to be a specific value, $14\text{mm}$, and for this minimum to occur 25mm into the opening of the
Figure 2.17: Final focus electrode.
Figure 2.18: Sheet beam gun in Solidworks®.
anode. Let $I$ denote the current of the beam. Next, we have two vectors $(x, z)_i$ and 
$(y, z)_i$, which describe the outer beam envelope in the $xz$ and $yz$ planes respectively. 
Note the $i$ subscript denotes the fact that we have this information at discrete locations 
along the $z$-axis. Let $y_j = \min(y_i)$ and the cathode anode spacing be denoted by $CAs$. 
Thus, our goal functions are

$$J_1 = \alpha_1(I - 11A)^2$$  \hspace{5cm} (2.14) 

$$J_2 = \alpha_2 \sum_i (x_i - 50mm)^2$$  \hspace{5cm} (2.15) 

$$J_3 = \alpha_3(y_j - 7mm)^2$$  \hspace{5cm} (2.16) 

$$J_4 = \alpha_4(CAs + 25mm - z_j)^2.$$  \hspace{5cm} (2.17) 

Our total goal function is given by

$$J = \sum_{i=1}^{4} J_i.$$  \hspace{5cm} (2.18) 

Our initial parameters are the cathode to anode spacing 150 mm, the cathode radius 7 
mm and the angle of the focus electrode $45^\circ$. Note that the weights were chosen such 
that $J_i(\lambda_0) = 1$. We used the implicit filtering algorithm for this problem. The final cost 
function values after this optimization are summarized in Table 2.3 and the performance 
goals are summarized in Table 2.4. The iteration history of the parameters and the goal 
functions are depicted in Figure 2.19. The beam shape in the $xz$ and $yz$ plane can be 
seen in Figure 2.20. We were able to come relatively close to all our goals and this design 
illustrates the potential to use our methodology with the BOA simulation code to design 
truly 3D electron devices.

<table>
<thead>
<tr>
<th>$\lambda_0$ (initial parameters)</th>
<th>$\lambda_{opt}$ (optimized parameters)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_1J_1$</td>
<td>1</td>
</tr>
<tr>
<td>$\alpha_2J_2$</td>
<td>1</td>
</tr>
<tr>
<td>$\alpha_3J_3$</td>
<td>1</td>
</tr>
<tr>
<td>$\alpha_4J_4$</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 2.3: Relative cost function values before and after optimization.
Figure 2.19: Iteration history of SBG optimization.

Figure 2.20: Particle trajectories in $xz$ and $yz$ planes, respectively.
Table 2.4: Beam characteristics achieved versus goals.

<table>
<thead>
<tr>
<th></th>
<th>$\lambda_{opt}$ (optimized parameters)</th>
<th>Goal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current</td>
<td>10.76A</td>
<td>11A</td>
</tr>
<tr>
<td>x width</td>
<td>87.4mm</td>
<td>100mm</td>
</tr>
<tr>
<td>$y_j$</td>
<td>12.9mm</td>
<td>14mm</td>
</tr>
<tr>
<td>$z_j$</td>
<td>24.46mm</td>
<td>25mm</td>
</tr>
</tbody>
</table>
Chapter 3

TWT Optimization

3.1 Introduction

Traveling Wave Tubes (TWTs) are vacuum devices invented in the early 1940’s (*U. S. Patent 2,3000,052. 1942, Kompfner 1976*) for amplification of radio frequency (RF) power. They are critical for radar, communications and electronic warfare missions of all Armed Services, as well as in certain commercial applications. Because of their high power, broad-bandwidth, compact size, and high efficiency, TWTs are used for satellite communications, airborne, ship borne, and ground-based radars, jamming, and decoy applications. Commercial applications include satellite communications, radar, and materials processing.

A basic schematic of a TWT is in Figure 3.1. The entire device is encased in a vacuum envelope. On the left end of the figure there is an electron gun similar to the ones described in Chapter 2. The TWT amplifies by converting the kinetic energy of an electron beam to the electromagnetic energy of an RF electromagnetic wave. The TWT amplifies a low power RF signal for transmission at significantly higher power. To maximize the energy transfer between the electron beam and the RF signal, the beam must flow in near synchronism with the RF wave, i.e., the electrons in the beam must travel at approximately the same velocity as the RF signal. Towards the output section of the TWT, the beam decelerates as energy is transferred to the RF wave. At the end of the beam tunnel the energy from the electron beam is then dissipated by the collector.
By decreasing the phase velocity of the RF signal, i.e. the velocity of the RF wave in the direction of the electron beam, as the beam decelerates, the RF wave will remain in synchronism for a longer distance, and efficiency will increase. This is typically achieved using a phase velocity taper, which varies the period (or other geometric parameter) of the slow-wave circuit in the output section of the TWT. In general, the phase velocity of the RF signal is proportional to the circuit period.

![Figure 3.1: Schematic of a TWT.](image)

Phase velocity tapers are also used to improve TWT linearity, and thus reduce distortion to enable high-data rate throughput (Kosmahl & Peterson 1984, Abe, Antonsen, Whaley & Danly 2002). For the specific application we describe here, several conditions must be met over a 3 GHz bandwidth centered at W-band (83.5 GHz). These are minimum small signal gain variation with frequency, minimum phase shift between small signal and saturated conditions, at least 10 W of output power, minimum variation in input power required for saturation at frequencies in the band, and maximum efficiency. There is a tradeoff between parameters, i.e., one must sacrifice phase shift to increase efficiency, for example. Thus, it is very useful for the designer to assign different relative weights to the various optimization goal functions, so that optimizations can focus on
the most important parameters for the specific application.

Most current design tools require manual input from the designer. Based on the simulation results, the designer makes judgements on parametric changes to more closely achieve the desired performance. These input modifications are made, and the simulation repeated. The process continues until the required performance is achieved or time and funding are exhausted. The primary cost is the intensive labor for the designer.

To avoid this tedious design procedure, optimization was incorporated into the physics-based helix TWT design code CHRISTINE-1D (Abe et al. 2002), as well as other TWT, large signal codes (Wilson 2001). To maximize the electronic efficiency, CHRISTINE-1D implements a phase velocity taper by varying the phase velocity of the slow-wave circuit at user-specified locations along the circuit. When the slow-wave circuit is physically altered to vary the phase velocity (such as by varying the period), the interaction impedance and attenuation will also change. The current version of CHRISTINE-1D assumes that the TWT incorporates a helix slow-wave circuit. With a helix, the phase velocity varies directly with the period, or pitch. Thus, if one reduces the pitch to 95% of its standard value, one can expect the phase velocity to be (approximately) 95% of its original value. With the current optimizer, the impedance is varied using the well-known sheath or tape helix model. These models use analytical functions to calculate the impedance as a function of helix geometry. Thus, the impedance is automatically calculated for a given helix geometry.

The first task of this research was to explore different optimization routines for optimizing TWT circuit taper designs. CHRISTINE-1D uses a modified steepest descent calculation in which the direction of travel in parameter space alternates between the steepest descent path and the average direction of travel during the recent cycles of minimization. It is well known that this type of gradient-based method, in general, performs very well when the optimization landscape is relatively smooth, and when a good 'first guess' for the solution can be made. However, for many important applications, including high power RF devices, the discrete nature of variables embedded in these problems coupled with the nonlinear interaction physics give rise to non-smoothness and non-convexity in the landscapes, which can defeat gradient-based methods. We propose using sampling methods, Nelder-Mead and DIRECT, which do not require gradient information but
rather sample the objective (goal) function on a stencil or pattern to determine the progress of the iteration and appropriately change the size, but not the shape, of the stencil. DIRECT is a global optimization routine with the ability to define a set of feasible design parameters. The routine will search this space for the global minimum. Since DIRECT is a global search method, it requires many iterations for high precision when the number of parameters is large. See Appendix A for an overview of these optimization methods. In this chapter, we explore using either DIRECT or Nelder-Mead algorithms, and DIRECT together with Nelder-Mead in a hybrid approach. That is, we initially apply the DIRECT method to generate a small set of iterates, then use Nelder-Mead to explore the smaller parameter space, thus accelerating convergence. This hybrid approach was previously demonstrated for designing Brillouin-focused electron guns by geometrically altering cathode and control surfaces to achieve specified beam characteristics (Lewis et al. 2004).

The second stage of this research was to apply the optimization methods with new goal functions to design a folded waveguide, slow-wave circuit. CHRISTINE-1D cannot model this geometry for optimization in its current form. More specifically, we cannot use the sheath or tape helix models to calculate variation of the impedance as a function of the phase velocity. Instead, it is computed using curve-fitting routines. In addition, since our design objectives for the folded waveguide slow-wave circuit are different than those built into CHRISTINE-1D, we propose several new goal functions.

Section 3.2 describes the application of sampling based optimization methods to a C-Band, helix TWT design problem, and their performance in comparison with the optimization algorithm in CHRISTINE-1D. Sections 3.3 and 3.4 contain our applications of the sampling methods for the optimal design of a folded waveguide, slow-wave circuit including the introduction of several new optimization goal functions.

3.2 Design Of A Linear, C-Band, Helix TWT

To compare these algorithms with the optimization routine in CHRISTINE-1D, we first optimized the same test case using CHRISTINE-1D and the Nelder-Mead algorithm. For our test case, we used the C-Band helix TWT design previously used by (Abe et al.
Table 3.1: Nominal electron gun parameters.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Nominal Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cathode Voltage</td>
<td>-4.80 [kV]</td>
</tr>
<tr>
<td>Beam Current</td>
<td>146 [mA]</td>
</tr>
<tr>
<td>Min. Beam Radius</td>
<td>0.158 [cm]</td>
</tr>
</tbody>
</table>

Table 3.2: Helix parameters common to all optimizations.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vacuum barrel radius</td>
<td>0.2794 [cm]</td>
</tr>
<tr>
<td>Helix radius</td>
<td>0.12446 [cm]</td>
</tr>
<tr>
<td>Helix tape width</td>
<td>0.03556 [cm]</td>
</tr>
<tr>
<td>Length of first stage</td>
<td>4.064 [cm]</td>
</tr>
<tr>
<td>Length of second stage</td>
<td>8.0776 [cm]</td>
</tr>
</tbody>
</table>

2002). The helix circuit parameters were designed around an existing, well-characterized, electron gun. This gun, designed by Northrop Grumman using a methodology described in (Whaley, Armstrong, Gannon, Groshart, Hurt, Hutchins & Roscoe 1998), produces a highly laminar beam with excellent transport properties, negligible helix interception at zero RF drive, and low harmonic growth. The nominal electron gun parameters are summarized in Table 3.1. The vacuum barrel radius, the helix radius, the helix tape width, and the rectangular cross-section BeO support rods were held invariant with respect to the different optimized designs. In this optimal design problem, only the helix pitch was modified according to a specified goal (or objective) function. The design was based on a two-stage circuit separated by a sever, located between $4.064 \leq z \leq 4.2164 [cm]$. Table 3.2 summarizes the common (invariant) parameters.

We wanted to optimize for peak power added efficiency, \( \eta_{eff} \), which is defined as the difference in RF output power, \( P_{out} \), and the RF input power, \( P_{in} \) divided by the beam power. This is given by the following goal function,

\[
\eta_{eff}(\lambda, P_{in}) = \frac{P_{out} - P_{in}}{I_b V_b},
\]

(3.1)

where \( \lambda \) represents the helix pitch taper, and \( I_b \) and \( V_b \) are the beam current and beam voltage, respectively. The optimal design problem is to vary the helix pitch as a function...
of distance to maximize the efficiency as defined by equation (3.1). Both optimization algorithms, CHRISTINE-1D optimization and the Nelder-Mead algorithm, used the same initial guess for the helix pitch, which was a constant pitch of 0.1\,[cm]/\,\text{turn} for all values. CHRISTINE-1D was executed using its native gradient based algorithm for 400 iterations, at which point all parameters were well converged, as shown in Figure 3.4. The optimizations were run on a 2.8GHz Toshiba laptop running Windows XP. This required approximately 110 minutes of CPU time, but as shown in Figure 3.2 after about 200 iterations the change in the helix pitch is negligible. Running CHRISTINE-1D using the Nelder-Mead algorithm required 686 iterations and 1051 function evaluations, which took approximately 26 minutes of CPU time. Figure 3.2 shows the convergence of the helix pitch values for the Nelder-Mead algorithm as a function of function evaluations. Note that the convergence for CHRISTINE-1D is with respect to iteration count while the convergence for Nelder-Mead is with respect to function evaluation. Nelder-Mead requires far less CPU time per iteration as it only requires a minimum number of function evaluations per iteration. See Appendix A for specific information on how an iteration of Nelder-Mead proceeds. The gradient information used by CHRISTINE-1D at each iteration generally requires more computational time to obtain than a few function evaluations. However, in the case of smooth objective functions gradient based algorithms can be far more effective per iteration. This implementation of Nelder-Mead uses the convergence criteria of small simplex size and small difference between function values on the simplex and function values of the current minimizer. CHRISTINE-1D uses a combination of user specified iterations and small step sizes in the parameter space. However, as this is a design problem computational time is not nearly as important as the ability to find a more efficient design, as long as the computational time is not prohibitive.

CHRISTINE-1D’s gradient optimization routine and the Nelder-Mead algorithm attained efficiencies of approximately 30% and 32% respectively. Figure 3.3 plots the electronic efficiency as a function of input power for the CHRISTINE-1D and Nelder-Mead optimized cases. From this test it appears that given a decent initial iterate both optimization routines return relatively similar results. However, small improvements in efficiency, as small as 1%, can have large financial impacts as illustrated in (Komm, Benton, Limburg, Menninger & Zhai 2001). This work describes a commercial satellite that
Table 3.3: Comparison of Nelder-Mead and steepest descent for random initial iterates.

<table>
<thead>
<tr>
<th>Opt. Routine</th>
<th>Case1</th>
<th>Case2</th>
<th>Case3</th>
<th>Case4</th>
<th>Case5</th>
<th>Case6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nelder-Mead</td>
<td>30.2%</td>
<td>26.0%</td>
<td>20.6%</td>
<td>20.0%</td>
<td>23.2%</td>
<td>31.9%</td>
</tr>
<tr>
<td>CHRISTINE-1D</td>
<td>30.0%</td>
<td>17.4%</td>
<td>25.3%</td>
<td>18.7%</td>
<td>21.7%</td>
<td>29.8%</td>
</tr>
</tbody>
</table>

has 10 kW of available power. This satellite can support 53 TWTs operating at 69% efficiency (the TWT and associated electronic power conditioner consume 187 W). A 1% improvement in the efficiency of each of the TWTs will allow the satellite to support an additional TWT. The average annual lease of a TWT is 2 million dollars per year and the average life span of a communication satellite is 15 years. Thus, a 1% improvement in efficiency could be potentially be worth 30 million dollars.

Another issue one must consider when working with an optimization routine is its robustness, i.e., how close must the initial iterate be to the minimum for the routine to converge to it. In the previous example, the initial iterate was close to optimum and both routines converged to a similar result. Now we would like to consider the case when the initial iterate is further from the minimum. We perturbed the initial helix values by multiplying each of the fifteen helix pitch values by $0.3v$ where $v$ is normally distributed with mean 1 and standard deviation 1, and ran both of the optimization routines. This experiment was repeated six times. The CHRISTINE-1D algorithm was run for 400 iterations for each optimization, which was sufficient for convergence in each case. The final peak efficiency is listed in Table 3.3. The average peak efficiency using the Nelder-Mead algorithm was 25.3% while the average peak efficiency for the CHRISTINE-1D algorithm is 23.8%. Thus, in conditions where a good initial iterate is available either routine may be successful. However, if finding a good initial iterate is difficult, then the Nelder-Mead algorithm may be a useful alternative. This may be particularly useful in the optimization of TWT multistage depressed collector stages where a good initial iterate is not always available.
Figure 3.2: Convergence of the helix pitch at fourteen axial grid locations from CHRISTINE-1D optimization as a function of iteration (each line represents pitch at a separate axial location). Convergence of the helix pitch at fourteen axial grid locations from the Nelder-Mead algorithm as a function of function evaluation (each line represents pitch at a separate axial location).
Figure 3.3: Comparison of CHRISTINE-1D’s native gradient-based optimization algorithm and the Nelder-Mead algorithm for maximizing the electronic efficiency.
Figure 3.4: Comparison of optimal helix pitches found by CHRISTINE-1D and Nelder-Mead.
3.3 Designing Folded Waveguide Slow-Wave Circuits

After performing this feasibility study on a known TWT design, we applied our optimization algorithms to the design of a folded waveguide, slow-wave circuit. This is a W-band, 9 kV, 28 mA TWT, which was tested in late-2006 as described in (Kory, Ives, Read, Booske, Jiang, van der Weide & Phillips 2005). CHRISTINE-1D was previously used to model a folded waveguide TWT, and the code gave good agreement with three-dimensional large signal analyses. This validated CHRISTINE-1D for modeling folded waveguide TWTs (Bhattacharjee, Booske, Kory, van der Weide, Limbach, Gallagher, Welter, Lopez, Gilgenbach, Ives, Read, Divan & Mancini 2004). Since we are using a folded waveguide circuit, not a helix circuit, we cannot use the CHRISTINE-1D internal sheath or tape helix models to calculate the impedance variation with phase velocity. Alternatively, we used the three-dimensional electromagnetic code Microwave Studio (*CST Microwave Studio Suite* 2005) to calculate the phase velocity and impedance as a function of the folded waveguide, slow-wave, circuit geometry. Using these results, we varied the impedance with phase velocity in the CHRISTINE-1D code, using a polynomial curve fit of the form

\[ f_z(i) = \alpha + \beta f_b(i) + \gamma f_b(i)^2 + \delta f_b(i)^3 + \epsilon f_b(i)^4, \]  

(3.2)

where \( f_z \) is the normalized impedance, and \( f_b \) is the normalized phase velocity at the \( i^{th} \) axial location along the circuit. The CHRISTINE-1D code does not currently account for the change in attenuation due to a change in helix pitch in the built-in optimizer.

The goal functions, which we use, require us to compute the performance of the folded waveguide TWT over ranges of both input power and frequency. The range of input powers was -25 [dBm] to 15 [dBm] with 30 total input powers where the model was evaluated. The frequency range was 82 [GHz] to 85 [GHz] with 5 total frequencies (82, 82.75, 83.5, 84.25, 85) where the model was evaluated. The inputs for our optimization were 25 values for the phase velocity. Using these values of phase velocity, the impedance was computed using the polynomial curve fitting function (3.2) and written to a CHRISTINE-1D input file by a MATLAB® function. We note that because our optimization routines, Nelder-Mead and DIRECT, are written in MATLAB®, the interface between CHRISTINE-1D and the optimization algorithms are handled through programs written in MATLAB®.
See Appendix A for a further description of these methods. CHRISTINE-1D was then run in the scanning mode with the input file computed as described above. Next, the output files from CHRISTINE-1D were read by MATLAB® to compute cost functions based on certain design specifications. We summarize this process in the flowchart depicted in Figure 3.5. The optimal design criteria are defined by goal functions. Our approach to treat the simulation and optimization codes as ”black boxes” allows us to easily implement different goal functions. We initially wanted to minimize the variation of the small-signal gain magnitude (the ratio of the output power to the input power at -25 dBm) with frequency. Therefore, we defined the following goal function

\[ \text{gaincost} = \sum_i (G_{\text{mag}}(-25\text{dBm}, i) - G_{\text{mag}}(-25\text{dBm}, 83.5\text{GHz}))^2, \quad (3.3) \]

where the gain magnitude \( G_{\text{mag}} = \frac{P_{\text{out}}}{P_{\text{in}}} \) is a function of the input power and frequency and the \( i \) index is taken over all frequencies not equal to 83.5 GHz. Next, we wanted to keep the phase change small, so our second goal function was

\[ \text{phasediff} = \sum_k \sum_j (\text{phase}(j, k) - \text{phase}(-25\text{dBm}, k))^2, \quad (3.4) \]

where the phase is a function of input power and frequency, the \( k \) index is taken over all frequencies of interest, and the \( j \) index is taken over all input powers of interest. We also wanted to make sure we achieved at least 10 W of output power at saturation at each frequency. However, no bonus was given for achieving more than 10 W of output power at saturation. Therefore, the desired goal function is

\[ \text{poutcost} = \sum_{P_{\text{out}, j}} (10W - P_{\text{out}}(P_{\text{in}}(i))), \quad (3.5) \]

where the value used is the output power at saturation for the \( i^{th} \) frequency and is only added if the output power falls below 10W. Note that if 10W of output power is achieved at all frequencies, this cost will be zero.

Next, we wanted saturation to occur at the same input power for each frequency. Let \( p_{\text{out}, j} \) denote the vector containing the 30 output powers corresponding to the 30 input powers with the \( j^{th} \) frequency. We then define \( I(j) \) to be the index in \( p_{\text{out}, j} \) where
Figure 3.5: Flowchart of folded waveguide optimization.

saturation occurs. The corresponding goal function for this design criterion is

$$\text{diffsat} = \sum_k |I(k) - I(83.5\text{GHz})|$$

where the $k$ index is taken over all frequencies not equal to 83.5 GHz. Finally, to maximize the efficiency, we define the following goal function

$$\text{effcost} = \frac{1}{\sum_i \sum_j \eta_{eff}(j, i)}$$

where the efficiency, $\eta_{eff}$, is a function of the input power and frequency and the $i$ index is taken over all frequencies of interest and the $j$ index is taken over all power values of interest. Note that instead of optimizing for peak efficiency we chose to optimize for a sum of the efficiencies over both power and frequency ranges. The idea behind choosing this goal function is that it looks for a design that is generally efficient across a wide range of frequencies and powers instead of only attempting to maximize the efficiency at power saturation, a power value at which the device is not generally operated. However, the sum over powers in equation \((3.7)\) will be dominated by values near saturation. Each
of the goal functions (3.3-3.7) were equally weighted to magnitude one when all phase velocities took their nominal, untapered values. The total combined goal function is a weighted sum of each of these goal functions as

\[\text{Cost} = a \times \text{gaincost} + b \times \text{phasediff} + c \times \text{poutcost} + d \times \text{diffsat} + e \times \text{effcost} \quad (3.8)\]

where \((a, b, c, d, e)\) are user specified constants weighting the relative significance of each goal function (3.3-3.7).

### 3.4 Optimal Design Results

First, we chose various weightings of the goal function and executed DIRECT for 40 iterations. Table 3.4 presents the most promising results using various combinations of weighting coefficients. If a goal function weighting is not listed, it is assumed to be zero. The goal function value is a percentage of the goal function when it is computed with a constant phase velocity. More specifically, if \(\text{gaincost}\) is 0.5 then the value of the \(\text{gaincost}\) goal function is half that of what it was for the non-optimized constant phase velocity. For each case, DIRECT was given a budget of 10000 function evaluations, though, to complete the 40 iterations, it required approximately 7000 function evaluations. We found that if we applied a substantial weighting \((e)\) for the \(\text{effcost}\), no weighting for \(\text{poutcost}\) was required. In other words, appropriately weighting the efficiency goal function achieved 10 W of output power over the operating frequency band. Also, note that the 0 value for \(\text{poutcost}\) indicates we achieved the minimum power specification for each of the optimizations.

DIRECT is a global sampling algorithm; however, global convergence usually requires a large and exhaustive search over the problem domain. In particular, DIRECT may perform many functional evaluations without improvement to the minimal value of the objective function (Finkel 2003). To overcome this shortcoming in the local behavior of the DIRECT algorithm, we explored the implementation of DIRECT algorithm in a hybrid approach. That is, we first used the DIRECT method as a starting-point generator. We then used the extremum values found by DIRECT as initial iterates for a local optimization method, such as the Nelder-Mead algorithm. Table 3.5 summarizes
Table 3.4: Results of DIRECT optimization for various combinations of weighting coefficients.

<table>
<thead>
<tr>
<th>Weighting</th>
<th>gaincost</th>
<th>phasediff</th>
<th>poutcost</th>
<th>diffsat</th>
<th>effcost</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>e = 1</td>
<td>1.23</td>
<td>1.99</td>
<td>0</td>
<td>1.20</td>
<td>0.85</td>
<td>0.85</td>
</tr>
<tr>
<td>b = 1, e = 10</td>
<td>0.61</td>
<td>0.44</td>
<td>0</td>
<td>1.18</td>
<td>0.94</td>
<td>9.84</td>
</tr>
<tr>
<td>b = 1, e = 5</td>
<td>0.48</td>
<td>0.23</td>
<td>0</td>
<td>1.18</td>
<td>0.97</td>
<td>1.2</td>
</tr>
<tr>
<td>a = b = 1, e = 10</td>
<td>0.45</td>
<td>0.33</td>
<td>0</td>
<td>1.21</td>
<td>0.95</td>
<td>10.28</td>
</tr>
<tr>
<td>a = b = 1, e = 5</td>
<td>0.29</td>
<td>0.13</td>
<td>0</td>
<td>1.21</td>
<td>1.02</td>
<td>5.52</td>
</tr>
</tbody>
</table>

Table 3.5: Optimization results using the hybrid approach (DIRECT followed by Nelder-Mead).

<table>
<thead>
<tr>
<th>Weighting</th>
<th>gaincost</th>
<th>phasediff</th>
<th>poutcost</th>
<th>diffsat</th>
<th>effcost</th>
</tr>
</thead>
<tbody>
<tr>
<td>e = 1</td>
<td>1.23</td>
<td>2.00</td>
<td>0</td>
<td>1.20</td>
<td>0.84</td>
</tr>
<tr>
<td>b = 1, e = 10</td>
<td>0.60</td>
<td>0.43</td>
<td>0</td>
<td>1.18</td>
<td>0.94</td>
</tr>
<tr>
<td>b = 1, e = 5</td>
<td>0.48</td>
<td>0.22</td>
<td>0</td>
<td>1.18</td>
<td>0.97</td>
</tr>
<tr>
<td>a = b = 1, e = 10</td>
<td>0.44</td>
<td>0.32</td>
<td>0</td>
<td>1.24</td>
<td>0.95</td>
</tr>
<tr>
<td>a = b = 1, e = 5</td>
<td>0.31</td>
<td>0.14</td>
<td>0</td>
<td>1.20</td>
<td>1.00</td>
</tr>
</tbody>
</table>

our results for this hybrid approach. Even though some better results were observed, the improvements were not significant, indicating that DIRECT method might already have achieved a good minimum. In general, we found that minimizing the combined goal function given by equation (3.8) required compromises between the various design components. For the specific application of interest, the folded waveguide TWT, the case where \(a = b = 1, e = 5\) provided the best compromise.

Table 3.6 summarizes the results for the non-optimized case, where the circuit period is uniform as a function of axial distance (no taper) and the optimal design corresponding to \(a = b = 1, e = 5\) case. As a function of frequency across the operating band, the table compares the range of saturated RF efficiencies, the small signal gain variation, saturated output power and the phase difference between small-signal and saturation. We note the large improvement in performance compared to the circuit geometry before optimization.
Table 3.6: Frequency dependent range of saturated RF efficiencies, small signal gain variation, range of saturated output power and range of phase differences between small-signal and saturation. The non-optimized case, where the circuit period is uniform as a function of axial distance (no taper), and the optimized case, $a = b = 1, e = 5$, are compared.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>No Taper</td>
<td>3.8-4.1</td>
<td>9.1</td>
<td>7-9</td>
<td>36-41</td>
</tr>
<tr>
<td>$a = b = 1, e = 5$</td>
<td>5.4-6.4</td>
<td>7.3</td>
<td>10-11.3</td>
<td>22-35</td>
</tr>
</tbody>
</table>
Chapter 4

HIV Model Analysis

4.1 Introduction

A great deal of effort has gone into modelling the physiologic and immunologic response of the HIV in individuals. For excellent reviews of the various types of modelling attempts see, e.g., (Callaway & Perelson 2002, Culshaw 2004, Perelson & Nelson 1999). In our attempt to model the physiologic and immunologic response of HIV in individuals, we will consider a variation of the model proposed in (Callaway & Perelson 2002). It accounts for a variation of drug efficacy based on target cell type. Note that the differentiation between target cells is based on cell type as opposed to cell location or longevity. A second essential component of this model is the modelling of the body’s HIV specific immune response. This aspect of the model is based on a Michaelis-Menten nonlinearity saturation as proposed by (Bonhoeffer, Rembiszewski, Ortiz & Nixon 2000). This specific model has been studied in (Adams 2005, Adams, Banks, Kwon & Tran 2004, Banks, Kwon, Toivanen & Tran 2006).

Various mechanisms by which the virus is treated or controlled have been considered. These mechanisms include drug treatment, e.g., reverse transcriptase inhibitors (RTIs) and protease inhibitors (PIs), and the bodies own virus specific immune response, e.g., cytotoxic lymphocytes (CTLs). RTIs interfere with the viral RNA-to-DNA synthesis thus blocking the introduction of viral DNA into the cells genetic information. Without this step the cells will not produce new virus. PIs attack the virus at the stage at which in-
fected cells assemble new virions thus producing non-infectious virions. CTL’s role in the immune response is to lyse antigen displaying cells thus killing them. A good introduction to biological basis behind these mechanisms can be found in (Alcamo 2003). Along with modelling these defense mechanisms comes the question of whether it is possible to manipulate these mechanisms in such a way as to benefit infected individuals. The prevailing medical practice is to prescribe highly active antiretroviral therapy (HAART). HAART generally consists of a combination of RTIs and PIs. In recent years this practice has prolonged the lifespan for infected individuals, however the drug cost of this type of treatment is high and the side effects can be potentially severe. Therefore a potential problem to consider in this area is whether there exists a drug treatment schedule that can sustain a low viral load and a healthy immune system while minimizing the amount of drugs used. This suggests an optimal control approach to treatment scheduling. This problem has been considered by a number of researchers. References (Banks et al. 2006, Joshi 2002, Wodarz & Nowak 1999, Kirschner, Lenhart & Serbin 1997) offer an example of the existing work. Our approach to the optimal control problem is similar to that taken by (Adams et al. 2004), with a difference being an explicit compartment for the non-infectious virus.

However, this chapter is not focusing on the control process but rather a question we would like to consider is what happens to the model dynamics under an optimal control based treatment schedule. We will further explore the control of this system in subsequent chapters. First, we would like to understand what aspects of the model have the most influence on the body’s immune response and its relationship to the drug treatment schedule. This can be studied by understanding the model sensitivity or how system parameter variation can affect system performance or ”output”, see for example (Eslami 1994, Stanley & Stewart 2002).

In addition to studying the model dynamics under an optimal control based treatment strategy, we would also like to study the feasibility of using stochastic estimation techniques on this model. Even when a model has been validated against experimental data it is still only an approximation of the true system dynamics. Also the data contains noise and is only a partial measurement of all the model compartments. With these two considerations a question arises as to how to combine these two estimates of the true
system dynamics and their associated uncertainties to obtain an accurate estimate of the true behavior of the system. For this problem we will consider a stochastic estimator, specifically the EKF.

The organization of this chapter is as follows. Section 4.2 outlines the specific aspects of the HIV model. In Section 4.3, we address some theoretical considerations related to using this model. We then describe an optimal control methodology in Sections 4.4 and 4.5. Sections 4.6 and 4.7 then consider the sensitivity analysis, parameter estimation, and filtering of this model under an optimal control based treatment schedule, respectively.

### 4.2 HIV Model

This model captures some of the inherent characteristics of HIV infection that we would like to consider. In addition to modelling the body’s HIV specific immune response and differentiating between types of target cells, this model also accounts for multi-drug therapy, both RTI and PI. An important physiological aspect of this model is its ability to reproduce a low non-zero viral load in the presence of multi-drug therapy. This model also possesses multiple stable equilibria including a viral dominant equilibrium and an immune response dominant equilibrium as can be seen in (Adams et al. 2004).

The dynamics of our HIV model are described by the system of differential equations

\[
\begin{align*}
\dot{T}_1 &= \lambda_1 - d_1 T_1 - (1 - \epsilon_1)k_1 V_IT_1 \\
\dot{T}_2 &= \lambda_2 - d_2 T_2 - (1 - f\epsilon_1)k_2 V_IT_2 \\
\dot{T}^*_1 &= (1 - \epsilon_1)k_1 V_IT_1 - \delta T^*_1 - m_1 E T^*_1 \\
\dot{T}^*_2 &= (1 - \epsilon_1)k_2 V_IT_2 - \delta T^*_2 - m_2 E T^*_2 \\
\dot{V}_I &= (1 - \epsilon_2)N_T \delta(T^*_1 + T^*_2) - [c + (1 - \epsilon_1)\rho_1 k_1 T_1 + (1 - f\epsilon_1)\rho_2 k_2 T_2] V_I \\
\dot{V}_{NI} &= \epsilon_2 N_T \delta(T^*_1 + T^*_2) - c V_{NI} \\
\dot{E} &= \lambda_E + b_E \frac{T^*_1 + T^*_2}{T^*_1 + T^*_2 + K_b} - d_E \frac{T^*_1 + T^*_2}{T^*_1 + T^*_2 + K_d} E - \delta_E E.
\end{align*}
\]

The state variables are $T_1$, the uninfected CD4+ T-cells; $T_2$, the uninfected target cells of a second kind; $T_1^*$, the infected T-cells; $T_2^*$, the infected target cells of a second kind; $V_I$, the infectious virus; $V_{NI}$, the non-infectious virus; and $E$, the immune effectors. We
Table 4.1: Values of parameters in the HIV model.

<table>
<thead>
<tr>
<th>parameter</th>
<th>value</th>
<th>unit</th>
<th>parameter</th>
<th>value</th>
<th>unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_1$</td>
<td>10.0</td>
<td>cells/mm$^3$/day</td>
<td>$\lambda_2$</td>
<td>31.98e-3</td>
<td>cells/mm$^3$/day</td>
</tr>
<tr>
<td>$d_1$</td>
<td>0.01</td>
<td>$1$/day</td>
<td>$d_2$</td>
<td>0.01</td>
<td>$1$/day</td>
</tr>
<tr>
<td>$k_1$</td>
<td>8e-4</td>
<td>virions/mm$^3$</td>
<td>$k_2$</td>
<td>0.1</td>
<td>virions/mm$^3$</td>
</tr>
<tr>
<td>$m_1$</td>
<td>0.01</td>
<td>virions/day/cells</td>
<td>$m_2$</td>
<td>0.01</td>
<td>virions/day/cells</td>
</tr>
<tr>
<td>$\rho_1$</td>
<td>1</td>
<td>cells</td>
<td>$\rho_2$</td>
<td>1</td>
<td>cells</td>
</tr>
<tr>
<td>$\delta$</td>
<td>0.7</td>
<td>$1$/day</td>
<td>$c$</td>
<td>13</td>
<td>virions/day</td>
</tr>
<tr>
<td>$f$</td>
<td>0.34</td>
<td>-</td>
<td>$N_T$</td>
<td>100</td>
<td>virions</td>
</tr>
<tr>
<td>$\lambda_E$</td>
<td>1e-3</td>
<td>cells/mm$^3$/day</td>
<td>$\delta_E$</td>
<td>0.1</td>
<td>$1$/day</td>
</tr>
<tr>
<td>$b_E$</td>
<td>0.3</td>
<td>$1$/day</td>
<td>$d_E$</td>
<td>0.25</td>
<td>$1$/day</td>
</tr>
<tr>
<td>$K_b$</td>
<td>0.1</td>
<td>cells/mm$^3$</td>
<td>$K_d$</td>
<td>0.5</td>
<td>cells/mm$^3$</td>
</tr>
</tbody>
</table>

do not give specific definitions of target cells of second kind. They could, for example be related to macrophages or brain cells. Table 4.1 contains values for model parameters. See (Adams, Banks, Davidian, Kwon, Tran, Wynne & Rosenberg 2005, Adams et al. 2004) for a further description of the model and model parameters.

The model also includes terms that model drug efficacy. $\epsilon_1$ is the efficacy of a RTI and $\epsilon_2$ is the efficacy of a PI.

### 4.3 Theoretical Considerations for Solutions to Model Equations

In this section we wish to discuss the existence and uniqueness of solutions to our model equations and then show that solutions to this equation are non-negative when initial conditions are non-negative. As we will show, we can only guarantee solutions to our optimal control problem will be integrable, thus we will be considering a differential equation, $\dot{x} = f(t, x)$, with a right-hand side which is continuous in $x$ but discontinuous in $t$. These are generally referred to as Carathéodory differential equations (Hale 1980, Filippov 1988). The following are the Carathéodory conditions used in the theoretical considerations of these equations. In the domain $D$ of the $(t, x)$-space, let $f$ be measurable
in \( t \) for each fixed \( x \), continuous in \( x \) for each fixed \( t \) and for each compact subset \( R \) of \( D \), there be an integrable function \( m_R(t) \) such that

\[
|f(t, x)| \leq m_R(t),
\]

for \((t, x) \in R\).

First, as the only time dependent terms in \( f \) are \( \epsilon_i(t) \) which we will show in Section 4.4 are integrable, \( f \) is measurable with respect to \( t \) for fixed \( x \). We will show that the model equations at least when replaced by saturated dynamics, as described below, are Lipschitz continuous. Since the treatment efficacies are bound between 0 and 1, equation (4.2) holds on compact sets. Thus, our model equations satisfy the Carathéodory conditions.

The primary property we will have to verify is that the model equations are Lipschitz continuous with respect to \( x \). As discussed in (Adams 2005) the general problem with appealing to Lipschitz continuity of the model equations, i.e., the \( f(t, x) \) term in \( \dot{x} = f(t, x) \), lie in the product nonlinearities such as \( k_1 T_1 V_1 \). The solution is to replace these terms with piecewise differentiable terms that saturate once viral or cell populations reach a certain level. Mathematically, this means introducing a saturated model \( \dot{x}^s = f^s(t, x^s) \) where terms such as \( ax_i x_j \), where \( a \) is a model parameter, are replaced with the term \( a^s_i(x_i) a^s_j(x_j) \), where

\[
a^s_m(x_m) = \begin{cases} 
0 & \text{if } x_m < 0, \\
\sqrt{a} x_m & \text{if } 0 \leq x_m \leq x^M_m, \\
\sqrt{a} x^M_m & \text{if } x^M_m < x_m.
\end{cases}
\]

(4.3)

Thus in the model \( \dot{x}^s = f^s(t, x^s) \) any term of the form \( a^s_i(x_i) a^s_j(x_j) \) can be bounded by \( a x^M_i x^M_j \). Also whenever our model states fall below the saturation limit, the saturated model equals the original model, i.e., \( f^s(t, x) = f(t, x) \).

To complete this proof we follow (Adams 2005) and separate the model equations into source, linear and nonlinear parts, i.e,

\[
f(t, x) = S + L(t)x + h(t, x).
\]

(4.4)

Note the saturation dynamics only affect the nonlinear term \( h \). Now, there is only one other type of nonlinear term in the original model and it is of the form \( b_{F^*+K} E \), where
\( T^* = T_1^* + T_2^* \). Note that this can be bounded by
\[
\frac{b}{T^* + K} E \leq bE, \tag{4.5}
\]
a linear equation in \( E \). The derivative of this term with respect to \( T^* \) can be bounded as well,
\[
\frac{bK}{(T^* + K)^2} E \leq \frac{bE}{K}. \tag{4.6}
\]
Note that this step requires that \( T^* \geq 0 \). We will show this holds as long as initial conditions for this model are non-negative. These bounds thus show that the derivative of the saturated nonlinear term is bounded above, i.e.
\[
\| D_x h^s(t, x) \|_\infty < \infty. \tag{4.7}
\]
Thus the Lipschitz continuity of the saturated dynamics can be shown.
\[
| f^s(t, x) - f^s(t, y) | = | L(t)(x - y) + h^s(t, x) - h^s(t, y) |
\]
\[
= | L(t)(x - y) + \int_0^1 D_x h^s(t, y + r(x - y))(x - y) dr |
\]
\[
\leq | L(t) | | x - y | + || D_x h^s ||_\infty | x - y | \tag{4.10}
\]
\[
= K_L | x - y |. \tag{4.11}
\]
Note that \( L(t) \) is bounded as the only time dependent term in this function is \( \epsilon_i(t) \) which is bounded. Also note that equation (4.9) is a consequence of a version of the fundamental theorem of calculus which is formally stated as Theorem 4.0.1 in (Kelley 1995). Thus our saturated model equations are Lipschitz continuous. Hence, the existence and uniqueness of our saturated model equation is guaranteed by the following theorem from (Hale 1980).

**Theorem 4.3.1** Let \( D \subset \mathbb{R}^{n+1} \) be an open set with an element of \( D \) written as \( (t, x) \), and \( f : D \to \mathbb{R}^n \) satisfies the Carathéodory conditions in \( D \). If for each compact subset \( R \) in \( D \) there exists an integrable function \( k_R(t) \) such that
\[
| f(t, x) - f(t, y) | \leq k_R(t) | x - y |
\]
for all \( (t, x) \in R \) and \( (t, y) \in R \), then for any \( (t_0, x_0) \in R \) there exists a unique absolutely continuous solution that passes through \( (t_0, x_0) \) and satisfies \( \dot{x} = f(t, x) \) almost everywhere.
Note that equation (4.12) holds because of the Lipshitz condition on $f$. Note that the above argument holds for $D = \mathbb{R}^{n+1}$, i.e., the saturated model dynamics are globally Lipschitz so we can extend the above theorem to show that model solutions exist and are unique for any finite time interval $I$. Since the solutions exist and are unique on any finite time interval $I$ they must be bounded on $I$.

As a note in our computations we never encountered any problems with unbounded growth or other numerical difficulties, so we were never forced to use the saturated dynamic equations.

In addition to the existence and uniqueness of solutions to the model equations we would also like to consider the continuity of the solutions with respect to variation in model parameters and initial data. We will use a theorem from (Filippov 1988).

**Theorem 4.3.2** Let the conditions of Theorem 4.3.1 hold. Let $f$ be continuous with respect to $q \in Q$ for almost all $t$. Let the solution to $\dot{x} = f(t, x, q)$, $x(t_0) = a$, $q = q_0$ be $\xi_0(t)$. Then the solutions to our model equations exist and are unique for $a$ and $q$ sufficiently near $a_0$ and $q_0$ and these solutions converge uniformly to $\xi_0(t)$ as $a \to a_0$ and $q \to q_0$.

The only additional condition we need to consider is the continuity with respect to $q$ and this is easily verified from the model equations (4.1).

As this model is biologically motivated another issue that we must consider is whether solutions are positive when initial conditions are positive. If this does not hold then our model could predict negative biological quantities. This trait should cause us to doubt the validity of such a model. In light of this we will make the following proposition.

**Proposition 1** Solutions to equation (4.1), with initial data $x(t_0) \geq 0$, have solutions $x(t) \geq 0$ for any time interval $I$.

**Proof** We can establish this result for the healthy target cells, $T_i$, and the immune effectors, $E_i$, by showing that the derivative of each state, $\dot{x}_i$, is independent of the other state variables and $\dot{x}_i \geq 0$ when $x_i = 0$. The equations for the healthy target cells $T_i$, $i = 1, 2$, are given by

$$\dot{T}_i(T_i = 0) = \lambda_i > 0$$  \hspace{1cm} (4.13)
when $T_i = 0$. The equation for the immune effectors $E$ are given by

$$\dot{E}(E = 0) = \lambda E > 0$$  \hspace{1cm} (4.14)

when $E = 0$. Thus solutions for $T_i$ and $E$ are non-negative if they are non-negative initially. The dynamics of $T_i^*$ and $V_I$ are not independent of other states when they are equal to zero and thus we must consider them by cases. If $T_i^* = 0$ and $V_I = 0$, then

$$\dot{T}_i^*(V_I = 0) = (1 - \epsilon_2)N_T\delta(T_1^* + T_2^*) > 0. \hspace{1cm} (4.15)$$

Note that this argument holds if only $T_1^* > 0$ or $T_2^* > 0$. If $T_i = 0$ and $V_I > 0$ then

$$\dot{T}_i^*(T_i^* = 0) = \alpha_i V_I T_i \geq 0,$$  \hspace{1cm} (4.16)

where $\alpha_i$ is a positive value. Note that inequality (4.16) follows from inequalities (4.13) and (4.15). Thus we have established that if $T_i^* = 0$ then $\dot{T}_i^* \geq 0$ and if $V_I = 0$ then $\dot{V}_I \geq 0$. Finally, if $V_{NI} = 0$ then

$$\dot{V}_{NI} = \epsilon_2 N_T \delta(T_1^* + T_2^*) \geq 0. \hspace{1cm} (4.17)$$

Thus, if any component of the model is zero then the derivative of that component of the model is non-negative. Hence, solutions must be non-negative when using non-negative initial conditions.

### 4.4 Optimal Control: HIV Model

In this section, we will consider drug treatment regimens using ideas from optimal control. We will mimic the approach taken by (Adams et al. 2004) and attempt to find a treatment regimen that minimizes the objective function

$$J(\epsilon_1, \epsilon_2) = \int_{t_0}^{T} [Q_v V(t) + R_1 \epsilon_1^2(t) + R_2 \epsilon_2^2(t) - SE(t)]dt. \hspace{1cm} (4.18)$$

Note that $\epsilon_1$ is bounded between 0 and $\epsilon_{1,\text{max}}$ and $\epsilon_2$ is bound between 0 and $\epsilon_{2,\text{max}}$. Also, note that if we write $x = [T_1 T_2 T_1^* T_2^* V V_{NI} E]^T$, $u = [\epsilon_1 \epsilon_2]^T$, and

$$J(u) = \int_{t_0}^{T} L(x, u, t)dt, \hspace{1cm} (4.19)$$
Then we can write the Hamiltonian for this problem as

$$H(x, u, p, t) = L(x, u, t) + p^T f(x, u, t),$$  \hspace{1cm} (4.20)

where \( p \) is a costate variable which arises from the use of Lagrange multipliers as this is really a constrained optimization problem. It can then be shown (Lewis 1995) that the solution to this problem is the control \( u^* \) (note the state \( x \) can be thought of as a function of \( u \), since once \( u \) is defined so is \( x \)) that satisfies the conditions

$$\dot{x} = \frac{\partial H}{\partial p} = f$$

$$-\dot{p} = \frac{\partial H}{\partial x} = \frac{\partial f^T}{\partial x} p + \frac{\partial L}{\partial x}$$  \hspace{1cm} (4.21)

and Pontryagin’s Minimum Principle

$$H(x^*, u^*, p^*, t) \leq H(x^*, u, p^*, t)$$  \hspace{1cm} (4.22)

where * denotes optimal quantities and the values of \( u \) are taken over all admissible values. Note that condition (4.22) implies that the Hamiltonian, \( H \), is to be minimized with respect to the control variable \( u \). For this reason the derivative of interest in our optimization routines are \( \frac{\partial H}{\partial u} \) or \( H_u \). For the specific optimality system for this model see Appendix [C].

Before attempting to compute an optimal control we will show that our problem has a solution at least with respect to the saturated dynamics as described in Section 4.3. We will use a result from (Fleming & Rishel 1975) to get the existence and uniqueness of our optimal control \( u^* \). Combining Theorem 3.4.1 and Corollary 4.1 we see that we need several conditions on our optimal control problem. They are as follows.

1. \( f \) is continuous and Lipschitz continuous with respect to \( x \).
2. \( L \) is continuous and convex on \( U \), for all \( t \in \mathbb{R} \) and \( x \in \mathbb{R}^n \), where \( U \) is the set of admissible controls.
3. \( U \) is closed and convex.
4. \( |f(t, x, u)| \leq c(1 + |x|) \).
5. \( L \geq c_1 ||u||^\beta - c_2, \ c_1, c_2 > 0, \ \beta > 1. \)

We established in Section 4.3 that our model equations at least when replaced with the saturated dynamics are Lipschitz continuous. Our admissible control set is \( U = \{(\epsilon_1, \epsilon_2) | 0 \leq \epsilon_1 \leq \epsilon_{1,\text{max}}, 0 \leq \epsilon_2 \leq \epsilon_{2,\text{max}}\} \), which is both convex and closed. Second, the integrand \( L \) is quadratic in \( u \) and thus convex on \( U \). Condition 4 holds for the saturated model because the nonlinearities are bounded, as can be seen in equations (4.3) and (4.5). Finally, we will establish condition 5

\[
L = q_1 V + r_1 \epsilon_1^2 + r_2 \epsilon_2^2 - SE \\
\geq \min\{r_1, r_2\} ||u||^2 + q_1 V - SE \\
\geq \min\{r_1, r_2\} ||u||^2 - SE \\
\geq \min\{r_1, r_2\} ||u||^2 - S_{\max}\{E\}.
\]

Note that the progression in the inequality to equations (4.25) and (4.26) are based on the previously established proposition that solutions to the model with non-negative initial conditions produce non-negative solutions and that solutions are bounded on any finite time interval. Thus condition 5 holds with \( c_1 = \min\{r_1, r_2\}, \ c_2 = \max\{E\} \) and \( \beta = 2 \) and we have established the existence of an optimal solution for our saturated model. Note that this result only guarantees that our optimal control is integrable, i.e. \( u^* \in L^1([t_0, T]) \).

## 4.5 Numerical Optimal Control

As closed form solutions to optimal control problem are available for only the simplest of model equations \( f \), and cost functions \( J \), we will employ a numerical approach to solving this problem. In (Sargent 2000), three general methods for numerically solving optimal control problems are described: solve the two-point boundary value problem as given by the necessary conditions, completely discretize the necessary conditions and solve these equations as a nonlinear program, or treat the problem as an optimization problem where computation of the cost and gradients depend on computation of the state and costate equations for a given control \( u \). As our admissible control set is bounded
and we must replace the stationary equation $H_u = 0$ with Pontryagin’s principle, the first method is not feasible. As our constraints only involve our control and they are relatively simple, $U = \{(\epsilon_1, \epsilon_2) | 0 \leq \epsilon_1 \leq \epsilon_{1,\text{max}}, 0 \leq \epsilon_2 \leq \epsilon_{2,\text{max}}\}$, it is more efficient to approach this problem as an optimization problem rather than a nonlinear program. See (Gunzburger 2003) for a further description of this method. We have chosen to use a quasi-Newton method to approximate the solution to our problem, as a steepest descent method can be slow to converge and the Hessian of the cost function can be difficult to calculate even in low dimensions as shown in (Kelley & Sachs 1987). Note that (Adams et al. 2004) introduces specific Lagrange multipliers to account for the bound constraint on the control. Rather than introduce these extra variables we account for this condition using a constrained optimization routine. It can be shown (e.g., see (Polak 1971)) that the first order approximation of the gradient of $J$, $\nabla J$ with respect to the $L^2$ inner product is given by

$$J(u + \delta u) - J(u) - \langle \nabla J, \delta u \rangle = o||\delta u||,$$  

where $\nabla J = H_u$. The process by which we compute an optimal control is as follows:

1. Given current control $u_c$, solve state equation forward in time for $x$
2. Given state and control, solve costate equation $p$ backwards in time
3. Using these quantities compute cost $J(u_c)$ and gradient $H_u$
4. Update control using optimization routine (projected BFGS)
5. Test for convergence, $||H_u|| < tol$
6. If stopping criteria is not met and maximum number of iterations not yet reached return to step 1

Using this cost and gradient information we employed a projected (input bound constrained to account for the limits of drug efficacy) BFGS code. The BFGS (Broyden, Fletcher, Goldfarb, Shanno) method is a quasi-Newton method which uses a rank-two update to approximate the true Hessian. See (Kelley & Sachs 1987) for an analysis of this routine applied to unconstrained optimal control problems. In addition, see Appendix A
for a brief description of the BFGS method. The code used for in this problem was written
by Tim Kelley and can be obtained from his home page at http://www.math.ncsu.edu/.
For more information on this optimization algorithm see (Kelley 1999).

Note that the initial condition for this problem is a healthy individual with a small
amount of free virus and infected cells introduced to their system. This can be termed
acute infection. The specific values for this initial condition are

\[ T_1 = 1000 \text{ cells/mm}^3, \quad T_2 = 3.198 \text{ cell/mm}^3, \]
\[ T_1^* = 10^{-5} \text{ cells/mm}^3, \quad T_2^* = 10^{-5} \text{ cell/mm}^3, \]  
\[ V_I = 0.001 \text{ virions/mm}^3, \quad V_{NI} = 10^{-5} \text{ virions/mm}^3, \quad E = 0.01 \text{ cells/mm}^3. \]  

Figure 4.1 illustrates the control computed by this process. Note that the optimal
control strategy calls for the patient to cycle on and off of the treatment. Figure 4.2
shows the model behavior under this sort of optimal control based treatment strategy.
Note that after the initial infection phase, where healthy T-cells drop and free virus levels
rise, the treatment reduces the free virus level and the number of healthy T-cells begin to
increase. During the ”drug holidays”, the period of treatment interruption the healthy
Figure 4.2: Model simulations under optimal control methodology.
T-cell count decreases and the viral load increases; however, at least initially the immune response is not stimulated. After several cycles of drug treatment and ”drug holiday”, the body’s own immune response begins to be stimulated. This optimal control methodology suggests that a structured treatment interruption (STI) type treatment strategy may be able to stimulate the bodies own immune response. There are several reasons to consider an STI type treatment strategy (Rosenberg, Davidian & Banks 2006). First, HAART can be complicated and have side effects. The ”drug holiday” offers a potential respite from the burden of continuous treatment. Secondly, as this optimal control methodology suggests, this type of treatment could potentially stimulate the body’s own virus specific immune response. An additional argument for treatment interruption occurs in individuals with drug-resistant strains. Biological evidence suggests that the mutations which result in drug-resistance may result in a loss in viral ”fitness” (Archer, Dykes, Gerondelis, Lloyd, Fay, Reichman, Bambara & Demeter 2000, Hance, Lemiale, Izopet, Lecossier, Joly, Massip, Mammano, Descamps, Brun-Vzinet & Clavel1 2001), e.g., the rate at which they are produced or infect target cells may be decreased. A treatment interruption would remove the environmental pressures which select the drug-resistant strain and allow the drug-sensitive strain to out compete the resistant strain. It is in the presence of this treatment that we will study the dynamics of this model.

4.6 Sensitivity Equations

Now that we have described the model and computed an optimal control based treatment strategy we would like to understand how the model behaves under this sort of treatment schedule. When one develops a mathematical model there are generally parameters that describe some physical/biological aspect of the model. For example, in the present HIV model there are parameters that describe such biological processes as cell birth and death rates. There are several situations in which one might want to understand the relationship between these model parameters and the behavior of the model. One situation would be if this parameter describes an unknown value and one would like to estimate this quantity from a set of experimental data, e.g., one has measurements of the T-cell count and the viral load and would like to estimate the rate at which free virus
infects T-cells from this data. Sensitivity analysis would determine which state outputs are sensitive to given parameters and whether one would expect to be able to estimate a parameter from a given set of data. Another situation would be the case where one would like to understand which parameters most influence certain components of the model. Sensitivity analysis could tell us whether the immune response is more dependent on the birth rate of the T-cells or the rate at which new virus is produced. This sort of analysis can help modellers to determine whether the given model matches the physical/biological laws governing a system. In both of these situations sensitivity equations can be useful.

Mathematically, the situation we want to consider is when the model is described by a system of ordinary differential equations. Thus the model has the form

$$\frac{dx(t)}{dt} = f(x, q),$$  \hspace{1cm} (4.29)

where $x \in \mathbb{R}^n$ denotes the state equations and $q \in \mathbb{R}^r$ denotes the parameters and the initial condition is $x(t_0) = x_0$. By sensitivity we mean how $x$ changes with respect to $q$, i.e., $\frac{\partial x(t)}{\partial q}$. Note that as $x$ is a function of time, so is the sensitivity equation. Now we can derive a system of differential equations for the sensitivities by differentiating both sides of equation (4.29) with respect to $q$ to get

$$\frac{\partial}{\partial q} \frac{dx(t)}{dt} = \frac{\partial f}{\partial x} \frac{\partial x}{\partial q} + \frac{\partial f}{\partial q}.$$  \hspace{1cm} (4.30)

Now if we switch the order of differentiation and couple this equation with equation (4.29) then we get a $n + nr$ dimensional system of differential equation for both the model and the sensitivities

$$\frac{dx(t)}{dt} = f(x, q),$$ \hspace{1cm} (4.31)

$$\frac{d}{dt} \frac{\partial x}{\partial q} = \frac{\partial f}{\partial x} \frac{\partial x}{\partial q} + \frac{\partial f}{\partial q}.$$ \hspace{1cm} (4.32)

Here, we assume $\frac{\partial x(0)}{\partial q} = 0$ as the initial conditions for the model would not be considered to be dependent on the parameters.

In addition to considering the sensitivities with respect to the parameters we would also like to consider the sensitivities with respect to the initial conditions, $\frac{\partial x}{\partial x_0}$. The theory and derivations of these sensitivities are analogous to the the sensitivities with
with respect to parameters so we will merely state that they solve the following initial value problem

\[
\frac{d}{dt} \frac{\partial x}{\partial x_0} = \frac{\partial f}{\partial x} \frac{\partial x}{\partial x_0} \tag{4.33}
\]

\[
\frac{\partial x(t_0)}{\partial x_0} = I, \tag{4.34}
\]

where \( I \) is the \( n \)-dimensional identity matrix. Note this initial condition comes from the fact that \( \frac{\partial (x(t_0))}{\partial (x_0)} = \delta_{ij} \), e.g., \( \frac{\partial E(t_0)}{\partial E_0} = 1 \) whereas \( \frac{\partial E(t_0)}{\partial x_0} = 0 \).

Often the expression for \( \frac{\partial f}{\partial x} \frac{\partial x}{\partial x_0} + \frac{\partial f}{\partial \delta} \frac{\partial \delta}{\partial x_0} \) would be derived analytically. Analytically computing the \( n + nr + n^2 \) sensitivity equations is prone to human error, e.g., this computation for the current problem would involve calculating 189 partial derivatives. Difference approximations can be used to calculate these derivative values; however, one must be careful to consider the relationship between the difference increment used to compute the derivative and the accuracy of the solutions obtained from the integrator. In this paper we compute \( \frac{\partial f}{\partial x} \) and \( \frac{\partial f}{\partial \delta} \) by using automatic differentiation (AD), see Appendix B. The specific code used to do AD in this work was written by Martin Fink and can be obtained at Matlab Central a file exchange for Matlab users.

Beyond traditional sensitivity analysis, generalized sensitivity functions (GSF) offer another way to analyze model dynamics (Thomaseth & Cobelli 1999, Batzel, Kappel, Schneditz & Tran 2006). Unlike traditional sensitivity analysis that offers information on the effect parameter variation has on model dynamics, GSF offers information on how changing data can affect parameter estimates. GSF also gives information on correlation between parameters and which portion of the data contains the most information for parameters estimation.

Consider discrete data \( z_k, k = 1, \ldots, M \) corrupted by noise which has zero mean and variance \( \sigma^2(t_k) \). Under this condition it can be shown that the change in optimal parameter estimate, \( \hat{q} \), which we will describe as optimal in the least squares sense, i.e.,

\[
\hat{q} = \arg\min_q \sum_{k=1}^{M} \frac{(x(t_k; q) - z(t_k))^2}{\sigma(t_k)^2}, \tag{4.35}
\]

when the data is changed is approximately (Thomaseth & Cobelli 1999)

\[
\delta \hat{q} \approx \left( \sum_{j=1}^{M} \frac{1}{\sigma^2(t_j)} \frac{\partial x}{\partial q} \frac{\partial x^T}{\partial q} \right)^{-1} \times \sum_{i=1}^{M} \frac{\partial x}{\sigma^2(t_i)} \delta z(t_i).
\] (4.36)

Given this approximation and the fact that \(\delta z(t_k) = \frac{\partial x(t_k)}{\partial q} \delta q\), we can then define the GSF which describes this variation during the course of the experiment

\[
GSF(t_k) = \sum_{i=1}^{k} \left( \sum_{j=1}^{M} \frac{1}{\sigma^2(t_j)} \frac{\partial x}{\partial q} \frac{\partial x^T}{\partial q} \right)^{-1} \times \frac{1}{\sigma^2(t_i)} \frac{\partial x}{\partial q} \frac{\partial x^T}{\partial q}.
\] (4.37)

Note that if we consider \(GSF(t) = 0\) for \(t < t_1\) then \(GSF(t)\) varies between 0 and 1 on the interval \((t_1, t_k)\). \(GSF(t)\) gives essentially two types of information about the model. First, we can determine which interval contains the most information about a given parameter by determining the interval in which \(GSF(t)\) changes most rapidly. Secondly, large oscillations of the \(GSF(t)\) can be indicative of parameter correlation.

### 4.6.1 Sensitivities: HIV

As one aspect of this work is to examine how well the EKF can be used to identify certain parameters from simulated data, we will rank the parameters by their relative sensitivity. It is important to look at the relative sensitivities, \(\frac{\partial x}{\partial q} \frac{\partial x}{\partial q}\). Merely looking at the derivatives, \(\frac{\partial x}{\partial q}\) can be misleading as Figure 4.3 indicates in the case of the parameter \(d_1\). These sensitivities are 4 orders of magnitude different. For the HIV model there are three types of potential data: total CD-4 count, viral load and immune effectors. Thus, we will rank the relative sensitivities with respect to these variables. For this analysis we will use the \(L^2\) norm to compare these sensitivity equations. Note this norm is defined as

\[
||f(t)||_2^2 = \int f^2(t)dt.
\] (4.38)
Since we are approximating the solution to our sensitivity equations $\frac{\partial x}{\partial q}$, we must also approximate this norm. For this ranking see Table 4.2.

![Comparison of CD4 sensitivity to $d_1$ between unscaled and scaled](image)

Figure 4.3: Difference between scaled and unscaled sensitivity for $d_1$.

One aspect of this analysis to note is the dominance of the sensitivities related to the equation for the immune effectors. In (Bortz & Nelson 2004) it is noted that this sort of relative sensitivity can be viewed as a logarithmic change, i.e., $\frac{q}{x} \frac{\partial x}{\partial q} = \frac{\partial \log(x)}{\partial \log(q)}$. Thus the relative sensitivities are dimensionless and can be thought of as a percentage change. The relative sensitivities of the parameters related to the immune response, e.g., $K_d$ and $b_E$, can have a magnitude of 100, as can be seen in Figure 4.4, which when viewed as a percentage change is quite large. It should also be noted that the sensitivity of these parameters is of a far different magnitude in the final hundred days than in the rest of the time interval. This sort of difference in sensitivity magnitude as a function of time is not generally seen in the parameters related to other compartments of the model. The difference in sensitivity of the parameters when the model is controlled by an optimal control based treatment strategy and when no treatment is used should also be noted.

As a comparison, the sensitivities with respect to the initial conditions are of a much smaller magnitude than the sensitivities with respect to the parameters and these sensitivities generally go to zero as time progresses. Figure 4.5 shows the sensitivity of the
Table 4.2: $L^2$ ranking and norm of sensitivities of parameters. The first number in each column is the ranking of each parameter with respect to the data, with 1 being the most sensitive and 20 being the least sensitive. The number in the parentheses is the actual $L^2$ norm. Note that the second horizontal line separates the parameters in the equation for the immune response.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$\frac{\partial T_1}{\partial q} + \frac{\partial T_1}{\partial q}$</th>
<th>$\frac{\partial V_I}{\partial q} + \frac{\partial V_{NI}}{\partial q}$</th>
<th>$\frac{\partial E}{E \partial q}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_1$</td>
<td>10 (99.506)</td>
<td>10 (164.2)</td>
<td>11 (7.8653)</td>
</tr>
<tr>
<td>$d_1$</td>
<td>11 (33.205)</td>
<td>11 (55.365)</td>
<td>12 (2.6285)</td>
</tr>
<tr>
<td>$k_1$</td>
<td>8 (110.56)</td>
<td>9 (170.88)</td>
<td>8 (19.332)</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>14 (13.305)</td>
<td>14 (27.842)</td>
<td>15 (0.47398)</td>
</tr>
<tr>
<td>$d_2$</td>
<td>18 (1.045)</td>
<td>18 (2.1067)</td>
<td>18 (0.03702)</td>
</tr>
<tr>
<td>$f$</td>
<td>19 (0.92713)</td>
<td>19 (1.9216)</td>
<td>17 (0.06691)</td>
</tr>
<tr>
<td>$k_2$</td>
<td>15 (7.0466)</td>
<td>15 (13.59)</td>
<td>16 (0.36099)</td>
</tr>
<tr>
<td>$\delta$</td>
<td>6 (155.38)</td>
<td>7 (249.86)</td>
<td>6 (58.937)</td>
</tr>
<tr>
<td>$m_1$</td>
<td>13 (14.951)</td>
<td>13 (28.453)</td>
<td>13 (0.75742)</td>
</tr>
<tr>
<td>$m_2$</td>
<td>17 (1.6183)</td>
<td>17 (3.4198)</td>
<td>19 (0.030954)</td>
</tr>
<tr>
<td>$N_T$</td>
<td>7 (113.57)</td>
<td>8 (171.13)</td>
<td>7 (19.798)</td>
</tr>
<tr>
<td>$c$</td>
<td>9 (103.82)</td>
<td>4 (507.13)</td>
<td>10 (17.032)</td>
</tr>
<tr>
<td>$\rho_1$</td>
<td>16 (2.9898)</td>
<td>16 (4.7365)</td>
<td>14 (0.61858)</td>
</tr>
<tr>
<td>$\rho_2$</td>
<td>20 (0.4526)</td>
<td>20 (0.8288)</td>
<td>20 (0.020118)</td>
</tr>
<tr>
<td>$\lambda_E$</td>
<td>12 (15.354)</td>
<td>12 (29.517)</td>
<td>9 (18.455)</td>
</tr>
<tr>
<td>$b_E$</td>
<td>1 (1252.9)</td>
<td>1 (2405.8)</td>
<td>1 (912.53)</td>
</tr>
<tr>
<td>$K_b$</td>
<td>5 (237.59)</td>
<td>6 (455.97)</td>
<td>5 (168.1)</td>
</tr>
<tr>
<td>$d_E$</td>
<td>2 (646.58)</td>
<td>2 (1241.9)</td>
<td>2 (478.7)</td>
</tr>
<tr>
<td>$K_d$</td>
<td>4 (253.47)</td>
<td>5 (486.48)</td>
<td>4 (180.44)</td>
</tr>
<tr>
<td>$\delta_E$</td>
<td>3 (524.15)</td>
<td>3 (1006.3)</td>
<td>3 (379.56)</td>
</tr>
</tbody>
</table>
CD4 cells with respect to the initial conditions. Note only the sensitivity with respect to $E_0$ does not go to zero in the controlled or uncontrolled case and the magnitude of this is very small in comparison to the sensitivities of the parameters related to this response.

As a note, the sensitivity rankings presented in (Adams 2005) are based on clinical treatment schedules and not those derived from an optimal control methodology. Those results do not show the same sort of sensitivity to the parameters in the equation for the immune effectors.

The sensitivity of the parameters related to the immune response is also seen in the generalized sensitivities as depicted in Figures 4.6 and 4.7. Figure 4.6 illustrates how changes in the immune effector $E$, especially, in the final hundred days, when this response is being stimulated, can have a large effect on the parameters that fit this response.

The GSF also indicates some extremely sensitive model dynamics when the immune effectors are stimulated (approximately the time interval 300 to 400 days) as can be seen in Figure 4.6. Note that all parameters indicate some oscillation in this region, which is indicative of parameter correlation. The parameters which were found to be most sensitive to the immune response $E$ in the classical sensitivity analysis have the largest amount of oscillation. This seems to indicate that the parameters most closely linked to the stimulation of the immune response are also quite sensitive to the data.

### 4.7 Extended Kalman Filter

Now that we have an understanding of model dynamics and parameter sensitivity under optimal control based treatment, we would like to look at the estimation and inverse problem. In light of the limits of deterministic modelling and the fact that we want to combine model prediction with physical data, we will explore the area of stochastic estimation. Chapter 1 of (Maybeck 1979) offers an excellent introduction to the basic ideas behind stochastic estimation and the Kalman filter. The example of finding one's location when lost at sea shows the basic concepts behind combining multiple measurements with varying levels of certainty about their accuracy (two people taking a star sighting to establish position with different levels of skill) and accounting for time
Figure 4.4: Relative sensitivities of parameters to total CD4 count, i.e., \( \frac{\partial T_1}{\partial q} \frac{q}{T_1} + \frac{\partial T_1^*}{\partial q} \frac{q}{T_1^*} \).
Figure 4.5: Relative sensitivities of initial conditions to total CD4 count, i.e., $\frac{\partial T_1}{\partial x_0} x_0 + \frac{\partial T_1^*}{\partial x_0} T_1^*$. 
Figure 4.6: Generalized sensitivities to $E$ under optimal control based treatment.
Figure 4.7: Generalized sensitivities to $T_1 + T_1^*$ under optimal control based treatment.
dynamics (the boat is moving). This example shows how the general stochastic estimation 
problem is to combine these two elements into a conditional distribution that describes 
the true state. We will provide a more rigorous treatment of how this distribution varies 
as the time dynamics influence the estimate in this research, but first we will introduce 
how one accounts for measurements. Following (Maybeck 1979), we assume that we have 
an estimate of our position, in one spatial dimension, \( x^- (t_1) \), that has some variance \( \sigma_x^2 \). 
Then, we obtain a measurement of the position \( z(t_1) \) that has variance \( \sigma_z^2 \). If we assume 
that both theses estimates are Gaussian distributions, then the best (in the maximum 
likelihood or least squares sense) combination of these two measurements is 

\[
x^+ (t_1) = \frac{\sigma_z^2}{\sigma_z^2 + \sigma_x^2} x^- (t_1) + \frac{\sigma_x^2}{\sigma_z^2 + \sigma_x^2} z(t_1) 
\]

\[
= x^- (t_1) + K (z_1 - x^- (t_1)) 
\]

where \( K = \frac{\sigma_x^2}{\sigma_z^2 + \sigma_x^2} \). The variance for our new estimate, \( \sigma^2 \), is \( \frac{1}{\sigma^2} = \frac{1}{\sigma_x^2} + \frac{1}{\sigma_z^2} \). Note, that we 
can see from these equations that if \( \sigma_z^2 \) is extremely large compared to \( \sigma_x^2 \), i.e., we have 
virtually no faith that this data is accurate compared to our current estimate, then we 
mostly disregard it and use \( x^- (t_1) \). The opposite is true if we completely trust the data 
and not our current estimate.

The general problem we want to consider is given a mathematical model for a physical 
system, \( x \), corrupted by some process noise, \( w \), and a model for data measurements, \( z \) 
also corrupted by some noise, \( v \), which is the best, in some sense, estimate for the true 
physical system. For this work we will consider the case when the model is continuous 
and the data measurements are taken at discrete times.

Mathematically, the problem can be stated as 

\[
\dot{x} = f(x, t) + g(t)w(t) \tag{4.41}
\]

\[
z_k = h(x(t_k), k) + v_k,
\]

where \( x(t_0) \) is normally distributed with mean \( \bar{x}_0 \) and covariance \( P_0 \), \( w(t) \) and \( v_k \) are 
white noise processes uncorrelated with \( x(t_0) \) and with each other and with means 0 and 
covariances \( Q \) and \( R \), respectively. The formulation of this problem in this manner was 
presented in (Lewis 1986). In the case where \( f \) and \( h \) are linear the problem reduces to 
the famed Kalman filter (KF) (Kalman 1960).
Note in this type of estimation we are trying to estimate statistics which describe a probability distribution on the true state of the process. In the KF the optimal estimate is defined by a normal distribution and thus only the mean, $\hat{x}$ and the covariance $P$ are needed to completely define the conditional distribution of the state dependent on the data, $p[x(t)|(z_k)]$. See (Welch & Bishop 2001) for an excellent introduction to the topic.

However, in the case where $f$ or $h$ are nonlinear the Kalman filter can no longer be employed and one must use some sort of approximation to the optimal estimator. As our measurements are linear and our model equations are nonlinear, in this section we will focus on how the optimal estimates evolve with respect to the time dynamics of the model, however there is a similar issue with respect to how the optimal estimate changes as measurements are taken. If we write $p = p[x(t)|z_k]$ then (Jazwinski 1970) shows how this distribution evolves according to the Fokker-Planck equation

$$\frac{\partial p}{\partial t} = -\text{trace} \frac{\partial [pf]}{\partial x} + \frac{1}{2} \sum_{i,j} \frac{\partial^2 [p(gQg^T)_{ij}]}{\partial x_i \partial x_j}, \tag{4.42}$$

where the notation $(\cdot)_{i,j}$ denotes the $ij^{th}$ entry in a matrix. This equation is generally not considered in applied estimation as analytical solutions are available in only the simplest of cases, e.g., in the one dimensional case if $f = 0$, $g = 1$ and $Q = 1$ this equation is the ordinary 1-D heat or diffusion equation

$$\frac{\partial p}{\partial t} = \frac{1}{2} \frac{\partial^2 p}{\partial x^2}. \tag{4.43}$$

As $f$ is nonlinear the distribution $p$ has no specific form. In order to study how this distribution evolves in time we will study how the first two moments, the conditional mean and covariance evolve.

First we will define the conditional expected value and the conditional covariance

$$\hat{x} = E[x(t)|(z_k; t_k \leq t)] \tag{4.44}$$

$$P(t) = E[(x - \hat{x})(x - \hat{x})^T|(z_k; t_k \leq t)]. \tag{4.45}$$

These are the two values the EKF will attempt to estimate. How these quantities evolve in time have been derived in (Lewis 1986) and (Jazwinski 1970) when no assumption is made about the function $f$ or the form of the conditional probability distributions $p$. We will include this derivation in expanded detail here.
Let $\phi : R^N \to R$ be such that $\phi \in C^2$. By the definition of expected value
\[
\hat{\phi}(x(t)) = \int \phi(x)pdx.
\] (4.46)

Note that we will use the notation $\hat{\cdot}$ or $E[\cdot]$ for expected value. Thus
\[
\frac{d\hat{\phi}(x(t))}{dt} = \int \phi(x) \frac{\partial p}{\partial t} dx.
\] (4.47)

Now accounting for equation (4.42) we get
\[
\dot{\hat{\phi}}(x(t)) = -\int \phi \text{trace} \frac{\partial [pf]}{\partial x} dx + \frac{1}{2} \int \phi \sum_{i,j} \frac{\partial^2 [p(gQg^T)_{ij}]}{\partial x_i \partial x_j} dx,
\] (4.48)

where $\dot{\hat{\phi}}$ denotes the time derivative. We will integrate this equation by parts in order to write this equation in terms of expected values. First, we will begin with the first integral in equation (4.48)
\[
\int \phi \text{trace} \frac{\partial [pf]}{\partial x} dx = \int \phi \sum_i (pf)_{xi}
\] (4.49)
\[
\quad = \sum_i \phi f_i |^\infty_{-\infty} - \int \sum_i \phi_{x_i} f_i p dx
\] (4.50)
\[
\quad = -\int \phi^T f pdx.
\] (4.51)

Note that this assumes that $p \to 0$ as $x \to \pm \infty$, a reasonable assumption for most probability density functions. Now, we will consider the second integral in equation (4.48)
\[
\int \phi \sum_{i,j} \frac{\partial^2 [p(gQg^T)_{ij}]}{\partial x_i \partial x_j} dx = \sum_{i,j} (gQg^T)_{ij} \int \phi_{x_i,x_j}
\] (4.52)
\[
\quad = \sum_{i,j} (gQg^T)_{ij} \left( \phi_{p_{x_j}} |^\infty_{-\infty} - \int \phi_{x_j} p_{x_j} dx \right)
\] (4.53)
\[
\quad = -\sum_{i,j} (gQg^T)_{ij} \left( \int \phi_{x_j} p_{x_j} dx \right)
\] (4.54)
\[
\quad = -\sum_{i,j} (gQg^T)_{ij} \left( \phi_{x_j} p |^\infty_{-\infty} - \int \phi_{x_i,x_j} pdx \right)
\] (4.55)
\[
\quad = \sum_{i,j} (gQg^T)_{ij} \left( \int \phi_{x_i,x_j} pdx \right)
\] (4.56)
\[
\quad = \int \text{trace}(gQg^T \phi_{xx}) pdx.
\] (4.57)
Now, combining equations (4.51) and (4.57) we get

$$\dot{\hat{\phi}}(x(t)) = \int \phi_x^T f p dx + \frac{1}{2} \int \text{trace}(gQg^T \phi_{xx}) p dx.$$  

(4.58)

This equation is, by definition,

$$\dot{\hat{\phi}}(x(t)) = E[\phi_x^T f] + \frac{1}{2} \text{trace} E[gQg^T \phi_{xx}].$$  

(4.59)

In order to get the time evolution equation for our conditional mean and covariance update we will use functions defined on components of $x$. To get the mean update, we let $\phi(x_i) = x_i$. Then $\phi_{x_i} = 1$ and $\phi_{x_i x_i} = 0$. Hence, equation (4.59) becomes

$$\frac{d\hat{x}_i}{dt} = E[f_i] = \hat{f}_i.$$  

(4.60)

To get the update for the covariance, one should first note that

$$P(t) = E[(x - \hat{x})(x - \hat{x})^T]$$  

(4.61)

$$= E[xx^T - x\hat{x}^T - \hat{x}x^T + \hat{x}\hat{x}^T]$$  

(4.62)

$$= E[xx^T] - \hat{x}\hat{x}^T - \hat{x}\hat{x}^T + \hat{x}\hat{x}^T$$  

(4.63)

$$= E[xx^T] - \hat{x}\hat{x}^T.$$  

(4.64)

Therefore

$$\frac{dP}{dt} = \frac{dE[xx^T]}{dt} - \frac{d(\hat{x}\hat{x}^T)}{dt}.$$  

(4.65)

Now consider the function $\phi(y) = x_i x_j$, where $y = [x_i x_j]$. Thus $\phi_y = [x_j x_i]^T$ and

$$\phi_{yy} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$  

(4.66)

Then, using equation (4.59) we obtain

$$\frac{d}{dt}E[x_i x_j] = E[(f_i x_j + x_i f_j) + \frac{1}{2} E[(gQg^T)_{ij} + (gQg^T)_{ji}]]$$  

(4.67)

$$= E[(f_i x_j + x_i f_j) + E[(gQg^T)_{ij}].$$  

(4.68)

Note that we have used that $gQg^T$ is symmetric. Now

$$\frac{d}{dt}(\hat{x}_i \hat{x}_j) = \hat{x}_i \frac{d}{dt} \hat{x}_j + \hat{x}_j \frac{d}{dt} \hat{x}_i.$$  

(4.69)
Using equation (4.60) we obtain
\[
\frac{d}{dt} (\dot{x}_i \dot{x}_j) = \dot{x}_i \dot{f}_j + \dot{x}_j \dot{f}_i. \tag{4.70}
\]
Now if we combine equations (4.65), (4.68) and (4.70) we obtain the equation for our covariance update
\[
\dot{P} = E[xf^T] - \dot{x}\dot{f}^T + E[fx^T] - \dot{f}\dot{x}^T + E[gQg^T]. \tag{4.71}
\]
Thus the equations that describe the time evolution of the first two moments of the conditional distribution are as follows
\[
\dot{\hat{x}} = \int f(x)p(x)dx \tag{4.72}
\]
\[
\dot{P} = E[xf^T] - \dot{x}\dot{f}^T + E[fx^T] - \dot{f}\dot{x}^T + E[gQg^T]. \tag{4.73}
\]
Note that these equations are not ordinary differential equations, and, in general, analytically unsolvable. The question then arises how can we approximate the solution to these equations. The idea of the EKF is to do a linear approximation of the nonlinear function at the current conditional mean, \(\hat{x}\), and use the simplified version of the above equations. Using this approximation we will show how this assumption, simplifies the time evolution equations. The linear approximation is
\[
f(x) = f(\hat{x}) + \nabla f(\hat{x})(x - \hat{x}). \tag{4.74}
\]
Now substituting this into the equation for the conditional mean update (4.72) gives
\[
\dot{\hat{x}} \approx \int (f(\hat{x}) + \nabla f(\hat{x})(x - \hat{x}))p(x)dx \tag{4.75}
\]
\[
= f(\hat{x}) \int p(x) + \nabla f(\hat{x}) \int xp(x)dx - \nabla f(\hat{x})\hat{x} \int p(x)dx \tag{4.76}
\]
\[
= f(\hat{x}) + \nabla f(\hat{x})\hat{x} - \nabla f(\hat{x})\hat{x}. \tag{4.77}
\]
\[
= f(\hat{x}). \tag{4.78}
\]
So we see that a first order approximation to the mean updates according to the deterministic dynamics. We also need a first order approximation for the expectation of the
nonlinear function $f$,

$$\hat{f}(x) = \int f(x)p(x)dx \quad (4.79)$$

$$\approx \int (f(\hat{x}) + \nabla f(\hat{x})(x - \hat{x}))p(x)dx \quad (4.80)$$

$$= f(\hat{x}). \quad (4.81)$$

Thus we observe that a first order approximation for the expectation of $f$ is $f$ evaluated at the conditional expectation. With those two facts, one can derive the equations for the conditional covariance update

$$\dot{P} = E[xf^T] - \hat{x}\hat{f}^T + E[f^Tx] - \hat{f}\hat{x}^T + E[gg^T]. \quad (4.82)$$

Now using the first order approximation for the expectation of $f$ and the fact that $g$ does not depend on $x$

$$\dot{P} = E[xf^T] - \hat{x}f(\hat{x})^T + E[f^Tx] - f(\hat{x})\hat{x}^T + gQg^T. \quad (4.83)$$

Now, we need to calculate two expectations

$$E[xf^T] = \int xf^T p(x|z_k)dx \quad (4.84)$$

$$\approx \int x(f(\hat{x}) + \nabla f(\hat{x})(x - \hat{x}))p(x)dx \quad (4.85)$$

$$= \hat{x}f(\hat{x})^T + E[x^T]\nabla f(\hat{x})^T - \hat{x}\hat{x}^T\nabla f(\hat{x})^T. \quad (4.86)$$

Similarly, one can obtain

$$E[f^Tx] = f(\hat{x})\hat{x}^T + \nabla f(\hat{x})E[xx^T] - \nabla f(\hat{x})\hat{x}\hat{x}^T. \quad (4.87)$$

Now, substituting these expectations into the equation for the conditional expectation

$$\dot{P} = \hat{x}f(\hat{x})^T + E[xx^T]\nabla f(\hat{x})^T - \hat{x}\hat{x}^T\nabla f(\hat{x})^T - \hat{x}f(\hat{x})^T$$

$$+ f(\hat{x})\hat{x}^T + \nabla f(\hat{x})E[xx^T] - \nabla f(\hat{x})\hat{x}\hat{x}^T - f(\hat{x})\hat{x}^T + gQg^T. \quad (4.88)$$

Simplifying this expression gives

$$\dot{P} = (E[xx^T] - \hat{x}\hat{x}^T)\nabla f(\hat{x})^T + \nabla f(\hat{x})(E[xx^T] - \hat{x}\hat{x}^T) \quad (4.89)$$

$$= P\nabla f(\hat{x})^T + \nabla f(\hat{x})P + gQg^T. \quad (4.90)$$
Thus we have our equations for the time update for the EKF.

The EKF then attempts to make an estimate of the true state with a "predictor-corrector" sort of implementation. First, in the "predictor" stage of the algorithm where no data is available and the dynamics are defined by the model, the following time update is employed

\[
\dot{\hat{x}} = f(\hat{x}, t) \tag{4.91}
\]

\[
\dot{P} = P\nabla f(\hat{x})^T + \nabla f(\hat{x}) P + gQg^T.
\]

These equations are integrated between time points \(t_k\) and \(t_{k+1}\). Then, when new data is available the "corrector" stage or measurement update is employed. This is the generalization of equation (4.40),

\[
K_k = P^{-1}(t_k)\nabla h(\hat{x})^T [\nabla h(\hat{x}) P^{-1}(t_k) \nabla h(\hat{x})^T + R]^{-1} \tag{4.92}
\]

\[
P(t_k) = [I - K_k \nabla h(\hat{x})] P^{-1}(t_k)
\]

\[
\hat{x}_k = \hat{x}_k^{-} + K_k [z_k - \nabla h(\hat{x}) \hat{x}_k^{-}].
\]

Note that this method only gives us the mean and covariance. Since \(f\) and \(h\) are nonlinear, the conditional distribution is generally non-normal and would require perhaps infinitely many statistics to fully describe it.

One of the general difficulties in employing this technique is the fact that the Jacobian of \(f\) and \(h\) must be computed. To overcome this difficulty, which is similar to that presented in the sensitivity study, we compute the Jacobians by AD.

As a note there are a variety of other approaches to estimate the propagation of the conditional distribution \(p[x|z_k]\). Early techniques include increasing the order the approximation of \(f\). An example of this technique is (Athans, Wishner & Bertolini 1968). The general result being increased accuracy of filter performance at the cost of increased computational burden.

In recent years alternate approaches to this problem have been introduced. These filters attempt to approximate the underlying conditional distribution by applying the nonlinear function to a set of points and recovering information on this distribution based on the effect the function has on these points rather than approximating the function.
itself like the EKF does. The Unscented Kalman Filter (Julier & Uhlmann 2004) and Gaussian filters (Ito & Xiong 2000) are examples of this type of filter. We attempted to use these filters, however states such as the viral load and infected cells achieve low levels in the presence of high drug treatment. The time-update of the mean and covariance in these filters are determined by sampling points in the neighborhood of the mean, \( \hat{x} \). The size of this neighborhood is determined by the current covariance matrix \( P \). Unless the covariance matrix was extremely small, elements in this sampling set had negative components. When initial conditions with negative elements (recall we showed that if the initial conditions are positive then solutions remained positive) were integrated through our model equations they gave unrealistic values. For this reason we used the EKF as our estimator.

4.7.1 Parameters Estimation and EKF

One further idea we would like to explore is the use of the EKF to not only estimate noisy and unmeasured states, but to also estimate unknown model parameters. Note that unlike the nonlinear least squares approach this technique does not require all of the data at once. This methodology only uses data as it is received and thus this routine can be used "online" to estimate parameters. This idea has been explored previously in (Ljung 1979), amongst others. As before we have a system that is parameter dependent and corrupted by noise, as well as errors in measurements. The system is described by

\[
\frac{dx(t)}{dt} = f(x, q) + g(t)w(t) \tag{4.93}
\]

\[
z_k = h(x(t_k), k) + v_k.
\]

In order to apply this technique we will augment our state equation and transform the unknown parameters into state variables with zero time dynamic, or written mathematically as

\[
\frac{dx(t)}{dt} = f(x, q) \tag{4.94}
\]

\[
\frac{dq(t)}{dt} = 0.
\]


Thus, in this form the EKF can be applied and the parameters will be estimated as if they were unknown states.

### 4.7.2 EKF: HIV

As a first test of the EKF we would like to use it in the context of an exact model (up to tolerance of the differential equation solver) with no parameter estimation. Our goal is two-fold: use the EKF as a state-estimator in the presence of incomplete noisy data and use the EKF to estimate certain model parameters in the process of the state estimation. We will implement the EKF methodology of equations (4.91) and (4.92) on the HIV mathematical model described by equation (4.1). The data we will use is as follows

\[
\begin{bmatrix}
T_1 + T_1^* \\
V_I + V_{NI} \\
E
\end{bmatrix}
+ r_k,
\]

where \( r_k \) is normally distributed with mean 0 and covariance \( R \). For the following simulations the following covariance \( R \) is used

\[
R = \begin{pmatrix}
4000 & 0 & 0 \\
0 & 4000 & 0 \\
0 & 0 & 10
\end{pmatrix}.
\]

The initial conditions for the model \( x_0^{\text{rand}} \) were randomly chosen. The \( i^{th} \) component of \( x_0^{\text{rand}} \) is given by \( (x_0^{\text{rand}})_i = |\gamma (x_0^{\text{true}})_i| \) where \( \gamma \) is normally distributed with mean 1 and variance \( \alpha \) and \( x_0^{\text{true}} \) is the initial condition used to create the simulated data. The absolute values were used to insure the initial conditions were positive as negative state values do not make physical sense. For the following simulations \( \alpha \) was set to 2.

Figure 4.8 shows the large amount of noise in the data. Note that some of the data is negative. While this is not physically realistic it does illustrate this method’s ability to filter poor data. It should be noted that when the filter predicted negative state values the actual prediction was set to 0. A similar strategy was employed in (Banks et al. 2006). Figure 4.9 shows the performance of the EKF as a state estimator. Note that \( V_{NI} \) was not included as once a virion becomes non-infectious it can no longer affect any
other model state. Also note that the EKF gives us an idea of a confidence interval for the state estimates in the covariance matrix $P$. For the plots in Figure [4.9] we included curves that are plus or minus 3 standard deviations of the true estimates if we ignore the covariance between states, i.e., $(x_{\pm})_i = \hat{x} \pm 3\sqrt{P}_{ii}$. The estimates track to the true states very quickly. As can be seen in Figure [4.10] the error from the perturbed initial conditions decreases rapidly in the first 50 days. The fact that we begin with an acute infection and the rapid state dynamics in this phase also account for some of the error in this time period. The other situation in which the error increases is during the treatment interruptions. This is when the model dynamics are changing most rapidly and several of the model states including $V_I$ are at their largest magnitude. In light of these two issues the filter seems to estimate all states of the model well.

![Graphs](image)

Figure 4.8: Filtering of noisy data, data taken daily. This filtering is the average over 20 runs. The plots on the right are the final 100 days of the filtering process.

### 4.7.3 Parameter Estimation and EKF: HIV

After having studied the EKF as a filter and state estimator, we will attempt to estimate some of the model parameters as well. The model parameters we will study are $N_T$, as it is sensitive to the data over the time course of interest, $d_2$, as it is relatively insensitive, and $b_E$, as it is sensitive when the immune response is stimulated. We will consider two different frequencies of data collection: the first being data is collected
Figure 4.9: Estimation of all states, data is taken daily. This filtering is the average over 20 runs. The plots on the right are of the final 100 days of the process.

Figure 4.10: Error between average estimate and true state, data taken daily. This estimation is the average over 20 runs. The plots on the left show the first 50 days of the process and the plots on the right show all of the simulation.
every day and the second being data is collected every five days. For the estimation
attempts where data is collected daily the same filter tuning parameters are used as in
the previous section, except for the choice of model error covariance \( Q \). As we are now
considering some model parameters as state variables, we must account for the fact that
while these parameters are inaccurate the discrepancies between the model prediction
and the the data cannot only be accounted for by observation noise. To overcome this
difficulty we will now assume \( Q \) to be time dependent, i.e., \( Q = Q(t_k) \). To get an idea of
the magnitude of \( Q(t_k) \) the model was simulated while perturbing the parameters that
we will be estimating with the EKF. For this experiment each of these parameters was
multiplied by a normal random variable with mean 1 and variance 0.1. Thus for each
perturbation we have a set of state vectors \( x_j \). Then a sample covariance was constructed

\[
Q(t_k) = \alpha \frac{1}{N} \sum_j^N (x_j(t_k) - \bar{x}(t_k))(x_j(t_k) - \bar{x}(t_k))^T,
\]

where \( \bar{x}(t_k) = \frac{1}{N} \sum_j^N x_j(t_k) \) and \( \alpha \) is a scaling factor to tune filter performance. For the
following simulations \( N = 8 \) and \( \alpha = 10^{-5} \).

The initial guesses for the parameters \( q_{0}^{\text{rand}} \) were randomly chosen. The \( i \)th component
of \( q_{0}^{\text{rand}} \) is given by \( (q_{0}^{\text{rand}})_{i} = |\gamma(q_{0}^{\text{true}})_{i}| \) where \( \gamma \) is normally distributed with mean 1
and variance \( \beta \) and \( q_{0}^{\text{true}} \) are the parameters used to create the simulated data. In the
following simulations \( \beta = 1 \). Note that in the parameter estimation problem the initial
conditions given for the state were the same as those used to create the simulated data,
i.e., \( x_{0}^{\text{rand}} = x_{0}^{\text{true}} \).

Figure 4.11 illustrates the ability of the EKF to filter noisy data even while estimating
unknown parameters. The average filter performance is not much different from Figure
4.8 when the model and model covariance were known exactly.

Figure 4.12 illustrates the performance of the EKF as a state estimator. All states
appear to be estimated relatively well except for \( T_2 \) and \( T_2^* \). This can perhaps be explained
by the fact that the data includes no element of either one of these state variables and
thus the EKF must rely largely on the model predictions to estimate these states. While
the model is inaccurate due to the parameter estimation one can expect the estimator to
struggle with state variables that are not particularly sensitive to the data. Figure 4.13
shows the $L^\infty$ error of the estimate of the states. Note that this estimation is not nearly as accurate as when the state model was exact, i.e. we were not attempting to estimate any parameters.

Figure 4.14 illustrates the estimation of each of the parameters. The performance reflects their sensitivity to the data. $N_T$ is sensitive to the data throughout the time course and is thus estimated relatively quickly and accurately. $d_2$ is relatively insensitive to the data and is thus estimated more slowly and with a larger confidence interval. $b_E$ is much more sensitive to the data as the immune response is stimulated and the estimate and confidence interval for the prediction of this parameter improves toward the end of the time course.

Figure 4.11: Filtering of noisy data, data taken daily. This filtering is the average over 20 runs. The plots on the right are the final 100 days of the filtering process.

Using the EKF on this model with data collection restricted to 5 days provided new challenges. In order to achieve consistent filter convergence it was necessary to remove some of the noise from the data and reduce the variation in initial guesses for parameter values. For this frequency of data collection the observation noise covariance matrix was
Figure 4.12: Estimation of all states, data taken daily. This filtering is the average over 20 runs. The plots on the right are for the final 100 days of the process.

Figure 4.13: Error between average estimate and true state, data taken daily. This estimation is the average over 20 runs. The plots on the right are the final 100 days of the process.
Figure 4.14: Estimation of parameters, data taken daily. This estimation is the average over 20 runs. The plots on the left illustrate the first 50 days of this process. The plots on the right show the parameter estimation over the entire time course.

chosen to be

$$R = \begin{pmatrix}
1000 & 0 & 0 \\
0 & 1000 & 0 \\
0 & 0 & 10
\end{pmatrix}. \quad (4.98)$$

Figure 4.15 illustrates the data used in this filtering attempt. This data contains the dynamics of the general time course for the CD-4 cells and the immune effectors, however the sparsity of the data generally only admits one measurement during the "drug holiday". This is why we did not consider taking data on any longer intervals between measurements.

Figure 4.16 illustrates the performance of the EKF as a state estimator. Once again $T_2$ and $T^*_2$ are the difficult state values to estimate. The estimate and confidence interval during the first 200 days are poor. However the performance improves in the final 200 days.

Figure 4.18 illustrates the performance of the EKF as a parameter estimation technique. As before $N_T$ is estimated relatively accurately, however it takes longer to improve
both the estimate and the confidence interval. $d_2$ appears to actually move away from the true parameter value. As this is a relatively insensitive parameter to the data this is not terribly surprising. As expected, the confidence interval for $b_E$ does not decrease substantially until the end of the time course when the immune response is stimulated.

Note the improved filter performance when the data is acquired daily instead of every five days. This raises the question of how well will this filter perform under clinical situations, when data may be more incomplete and may be taken more infrequently.

Figure 4.15: Filtering of noisy data, data taken daily. This estimation is the average over 20 runs.
Figure 4.16: Estimation of all states, data taken every 5 days. This estimation is the average over 20 runs.
Figure 4.17: Error of true states, data taken every 5 days. This estimation is the average over 20 runs.
Figure 4.18: Estimation of parameters, data taken every 5 days. This estimation is the average over 20 runs.
Chapter 5

HIV: RHC

5.1 Introduction

There have been various studies illustrating the application of optimal control methodology to HIV drug treatment. A large number of methodologies for the control of HIV are based on open-loop techniques, i.e., a control is pre-computed for a given model and initial conditions. An example of these works can be seen in (Kirschner et al. 1997, Joshi 2002, Adams et al. 2004, Culshaw, Ruan & Spiteri 2004, Fister, Lenhart & McNally 1998, Garira, Musekwa & Shiri 2005). This technique may be inadequate for a number of reasons. First, some unmodelled effect can disturb the system, thus rendering the treatment schedule ineffective or, even worse, detrimental. Secondly, if a patient misses a dose, the open loop control is no longer optimal or even necessarily beneficial. This methodology also does not take advantage of the periodic measurements taken from an individual. For these reasons we would like to consider feedback methodology, where the control depends on the current state of the system, in the context of the control of an HIV infection.

Some feedback techniques may be inappropriate or non-applicable to the problem of HIV control. Many feedback techniques are based on linear plants or models. It is difficult to accurately model HIV dynamics with a linear model, especially as the time scale expands and the number of modelled biological mechanisms increases. See (Perelson & Nelson 1999) for an excellent review of how the complexity of these models has increased
as understanding of disease dynamics has progressed. Techniques based around solutions of Ricatti equations may be augmented to solve nonlinear control problems; however, these techniques require that the model be of a certain form and can require frequent state knowledge or estimation. For an example of this sort of technique applied to HIV control see (Banks et al. 2006). A potential solution involves using Bellman’s principle of optimality, however this involves solving a nonlinear partial differential equation which may be computational difficult.

RHC seeks to gain the benefits of a feedback control while utilizing the computational simplicity of a calculus of variations approach. RHC solves a finite horizon open-loop optimal control problem on-line at each sampling instant. There are several design considerations in using this methodology including the process model, the cost function to be minimized, the sampling period, the control horizon, and the method by which the state is obtained at each sampling instant. The method by which the state is obtained at each sampling instant is of especial concern as it generally involves some combination of model prediction and noisy measurements. For this problem, we will employ stochastic estimation; in particular, the EKF. An excellent survey paper related to the theory behind RHC is (Mayne, Rawlings, Rao & Scokaert 2000). Another source for the multitude of industrial applications of this technique is (Qin & Badgewell 1997). This technique has been used several times in the context of HIV control (Shim, Han, Chung, Nam & Seo 2003, Zurakowski, Messina, Tuna & Teel 2004).

The contribution of this chapter is that it considers a complex model of HIV dynamics that includes multi-drug therapy, multiple target cells and a compartment accounting for the virus specific immune response. It uses RHC in conjunction with stochastic estimation to control HIV infection under poor drug adherence, noisy measurements, and inaccurate model parameters.

The organization of this chapter is as follows. We describe the specific RHC methodology that we have considered in Section 5.2. Numerical results for the RHC implementation are presented in Section 5.3. Situations in which treatments are missed and a state estimator is employed are presented in Sections 5.4 and 5.5 and the solution to these two combined problems is illustrated in Section 5.6. Section 5.7 shows how RHC methodology can be used in situations in which model parameters are allowed to vary in
some unknown manner.

## 5.2 RHC Methodology

As stated before, RHC seeks to gain the benefits of a feedback control while utilizing the computational simplicity of a calculus of variations approach. Thus, there are several basic components to the methodology. First, there are the model equations which we use to predict system behavior and compute the optimal control. There is the estimation of the current state, which can be done with a combination of measured data and predicted output. We will use the methodology described in Section 4.7, however there are a variety of techniques to complete this step. There is also the specific optimal control problem, or cost function one will seek to minimize. Finally, there is the methodology by which control horizon, the period over which each optimal control problem is solved, and the period over which this control is used is selected. A schematic that illustrates this general process is in Figure 5.1.

![Figure 5.1: Schematic diagram of RHC methodology.](image)

To make this methodology more concrete we will present the following mathematical formulation. We will use notation for the RHC methodology that is similar to that
presented in (Ito & Kunisch 2001). Let \([t_i, t_{i+1}]\) be a sequence of time intervals. Let \(t_{ch,i}\) such that \(t_{ch,i} \geq t_{i+1} - t_i\) be the control horizon. Consider the sequence of problems, \(P_i\),

\[
\min_u J(u) = \int_{t_i}^{t_{i+1}+t_{ch,i}} L(x, u, t) dt + \phi(x(t_i + t_{ch,i}))
\]

subject to

\[
\dot{x} = f(x, u), \quad x(t_i) = x_i,
\]

on the interval \([t_i, t_{i+1}]\). In order to be biologically realistic we will also enforce bounds on the control, \(u_{min} \leq u(t) \leq u_{max}\). Note that for this problem the control is the efficacy of the drug dose and thus we will write \(u = [\epsilon_1 \epsilon_2]\).

The solution to each optimal control problem \(P_i\) can be analyzed and solved in the manner presented in Section 4.4.

The methodology involved in using RHC is as follows:

1. Given initial condition \(x(t_i)\) solve the control problem \(P_i\) on the time interval \([t_i, t_i + t_{ch,i}]\).

2. Use the control as it is defined on the interval \([t_i, t_{i+1}]\) to determine the trajectory on the same time interval.

3. Use observations or an estimator to determine \(x(t_{i+1})\).

4. Repeat the process over starting at the next time interval \(t_{i+1}\).

This process is illustrated in Figure 5.2. Note for this example \(t_i = i\), \(t_{ch,i} = 10\). This means that \(P_0\) was solved by minimizing the integral from 0 to 10. However the control was only used from 0 to 1. At this point a new control was computed by solving \(P_1\) on the interval 1 to 11. This control was then used from 1 to 2. Note that neither the time intervals nor the control horizon need to be constant.

Note that our numerical routine to solve each control problem \(P_i\) is an iterative method thus requiring an initial iterate. The initial iterate for \(P_0\) was \(u_0 = 0.5[\epsilon_{1,max} \epsilon_{2,max}]\). The initial iterate for each subsequent \(P_i\) was the optimal solution to \(P_{i-1}\). Two heuristic approaches were used to facilitate numerical computations. One approach was how to choose the initial iterate when the immune response \(E\) was stimulated above 200
Figure 5.2: Example of RHC.
Table 5.1: Parameters for RHC problem.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_i$</td>
<td>50 days</td>
</tr>
<tr>
<td>$t_{ch,i}$</td>
<td>50 days</td>
</tr>
<tr>
<td>$Q_e$</td>
<td>$10^{-9}$</td>
</tr>
<tr>
<td>$R_1$</td>
<td>$10^{-9}$</td>
</tr>
<tr>
<td>$R_2$</td>
<td>$10^{-9}$</td>
</tr>
<tr>
<td>$S$</td>
<td>$10^{-5}$</td>
</tr>
<tr>
<td>$\theta$</td>
<td>$5^{-4}$</td>
</tr>
</tbody>
</table>

cells/mm$^3$. In this case, the initial iterate was decreased by a factor of 10 as this facilitated convergence to a low dose control that stimulated further immune response. The other was in the case that the ode solver did not converge for a candidate control vector. In this case $u_p$ was replaced with $0.99u_p$ and the algorithm proceeded.

5.3 Numerical Results: Perfect State Knowledge

For our study we have employed the following cost functional

$$J(\epsilon_1, \epsilon_2) = \int_{t_i}^{t_i+t_{ch,i}} \left[ Q_e V(t) + R_1 \epsilon_1^2(t) + R_2 \epsilon_2^2(t) - S E(t) \right] dt + \theta (E(t_i + t_{ch,i}) - 353.108)^2,$$

(5.3)

where $Q_e, R_1, R_2, S$ and $\theta$ are weighting coefficients. The maximum values for the drug efficacies for the following simulations are $\epsilon_{1,max} = 0.7$ and $\epsilon_{2,max} = 0.3$. The results of Section 4.4 show that each optimal control problem $P_i$ has a unique solution with the addition of one condition, that $\phi$ is continuous. As $\phi$ is a quadratic function of $E$ this condition holds.

Table 5.1 contains the values for the optimal control problem. The small values of weighting coefficients are used so that the gradient values are of approximately the same magnitude as the control bounds thus improving the efficiency of the optimization.

In this section we will assume that we have perfect knowledge of the state at each point in which a new control is computed. This is not physically realistic, but we will use
the EKF to address this problem in Section 5.5. The initial condition for this problem is a healthy individual with a small amount of free virus introduced to their system. This can be termed acute infection. The specific values for this initial condition are

\[
T_1 = 1000 \text{ cells/mm}^3, \quad T_2 = 3.198 \text{ cell/mm}^3, \\
T_1^* = 10^{-5} \text{ cells/mm}^3, \quad T_2^* = 10^{-5} \text{ cell/mm}^3, \\
V_I = 0.001 \text{ virions/mm}^3, \quad V_{NI} = 10^{-5} \text{ virions/mm}^3, \\
E = 0.01 \text{ cells/mm}^3.
\]  

(5.4)

Our goal is to transfer this infected individual to a healthy immune response dominated equilibrium. The specific values for this state are

\[
T_1 = 967.839 \text{ cells/mm}^3, \quad T_2 = 0.621 \text{ cell/mm}^3, \\
T_1^* = 0.076 \text{ cells/mm}^3, \quad T_2^* = 0.006 \text{ cell/mm}^3, \\
V_I = 0.415 \text{ virions/mm}^3, \quad V_{NI} = 0 \text{ virions/mm}^3, \\
E = 353.108 \text{ cells/mm}^3.
\]  

(5.5)

This equilibrium has been studied in (Adams et al. 2005, Adams et al. 2004). We will merely note here that this is a stable equilibrium in the presence of no treatment. We have studied three cases for how often the control is allowed to change. The treatments are allowed to vary daily, every five days, and every ten days. Note that as the model and control are continuous function (as opposed to a discrete model) the control values used in between specified treatment values are a linear interpolation. Figures 5.3, 5.4, 5.5, 5.6 show control and model response for each of these cases. The non-infectious virus \(V_{NI}\) has been omitted in the figures as that compartment has no direct effect on any of the other compartments. Figure 5.4 shows the phase plot of the log of the immune effector as a function of the log of the virus. This plot begins at acute infection and thus relatively low levels of \(E\) and \(V\). We can see how the viral load experiences one increase during infection and four increases during treatment interruptions. It is interesting to see that \(E\) decreases during most of the treatment interruption only to increase toward the end of the interruption. Note that each one of these control laws has treatment interruptions under which the viral load and infected cells rebound, while the number of healthy cells drops.
Also note that how often the control is allowed to vary has a large effect on how long it
takes the immune response to be stimulated and for the individual to be transferred to
the healthy steady state. When the control is allowed to vary daily the control is largely
turned off after 400 days, whereas it is not turned off until approximately 600 days when
the control varies every 5 days and 900 days when the control is allowed to vary every
10 days. The phase plots in Figure 5.4 also illustrate the difference in system response
when the control is only allowed to vary every 10 days. Note how more interruptions are
required and there is a good deal more time where the immune response is declining.

Figure 5.3: Model response and control when the control is allowed to vary daily.

We would also like to determine whether this methodology can be used to transfer
a patient from the unhealthy equilibrium to the healthy equilibrium. By unhealthy
equilibrium we mean the state

\[
T_1 = 163.573 \text{ cells/mm}^3, \quad T_2 = 0.005 \text{ cell/mm}^3, \\
T_1^* = 11.945 \text{ cells/mm}^3, \quad T_2^* = 0.046 \text{ cell/mm}^3, \\
V_I = 63.919 \text{ virions/mm}^3, \quad V_{NI} = 0 \text{ virions/mm}^3, \\
E = 0.024 \text{ cells/mm}^3.
\]  

(5.6)

This equilibrium represents a physical state in which the immune response has largely
been destroyed and a large quantity of infected cells and free virus is present in an
individual. Note that this equilibrium is stable. Figures 5.7, 5.9 and 5.10 illustrate the
Figure 5.4: Phase plot of log of immune effectors vs. log of virus when control is allowed to vary daily on the left. The results for when control was allowed to vary every 10 days are on the right.

Figure 5.5: Model response and control when the control is allowed to vary every 5 days.
model simulations and controls when the initial conditions are given by the unhealthy steady state. Figure 5.8 shows the phase plot of the log of the immune effector as a function of the log of the virus. In each one of these situations, even when the control was only allowed to vary every 10 days, the RHC methodology was able to transfer the system to the immune effector dominated equilibrium.

Figure 5.7: Model response and control when control is allowed to vary daily. Initial conditions given by unhealthy steady state.
Figure 5.8: Phase plot of log of immune effectors vs. log of virus when control is allowed to vary daily and the initial condition is the unhealthy equilibrium.

Figure 5.9: Model response and control when control is allowed to vary every 5 days. Initial conditions given by unhealthy steady state.
5.4 Unscheduled Treatment Interruption

A natural question to pose is what if a patient goes off medication? This can occur because of new infections, drug side effects, or a number of other reasons. Open loop methodology has no way to take this into account and will most likely prove insufficient. Feedback controllers can take this treatment lapse into account and adjust the controller. We will examine how RHC performs when a patient disregards the proposed treatment schedule and goes off treatment. For the following simulations a patient went off treatment for 10 days after 300 days of RHC based treatment. Other than this stopping of treatment the experimental design is identical to that presented in Section 5.3. Figure 5.11 contains the results from this simulation. Note that this situation requires 6 periods of treatment interruptions whereas the identical situation in which the treatment regiment was followed required only 4 treatment interruptions.

As the immune response is key to the healthy steady state we should examine the effect this treatment interruption has on the immune response. Figure 5.12 has a comparison of the immune response when the control is allowed to vary daily but in one case the patient adhered to the treatment protocol whereas in the other case the patient went off treatment. First it delays the stimulation of the immune response by approximately 120 days. However it should also be noted that the RHC based treatment strategy was able
Figure 5.11: Model response and control when control is allowed to vary daily and control was interrupted after 300 days for 10 days.

to overcome the failure to take medication and still transferred the patient to healthy steady state. Figure 5.13 repeats this situation when the control is allowed to vary every 5 days. Note that when the control can only vary every 5 days the stimulation of the immune response is delayed by approximately 180 days, a longer delay than when the control can vary daily.

Figure 5.12: Immune response when control is allowed to vary daily. Comparison between following treatment and when treatment was interrupted after 300 days for 10 days. Plot on the left contains the full time interval, while the plot on the right focuses on the time period directly after the patient went off medication.
Figure 5.13: Immune response when control is allowed to vary every 5 days. Comparison between following treatment and when treatment was interrupted after 300 days for 10 days. Plot on the left contains the full time interval, while the plot on the right focuses on the time period directly after the patient went off medication.

5.5 Numerical Results: State Estimation

In this section, we will address the problem of implementing RHC methodology while relying on the EKF to estimate the state each time a new control is implemented. We will use simulated data created by the model, corrupted by observation noise. The only process noise we will assume is based on the accuracy of the integrator and this is assumed small. Thus we will choose $Q = 1^{-6}I$. The form of the data we are going to use is

$$z_k = \begin{pmatrix} T_1 + T_1^* \\ V_I + V_{NI} \\ E \end{pmatrix} + r_k,$$

where $r_k$ is normally distributed with mean 0 and covariance $R$. For the following experiments $R$ will be of the form

$$R = \begin{pmatrix} r_1 & 0 & 0 \\ 0 & r_2 & 0 \\ 0 & 0 & r_3 \end{pmatrix},$$

where $r_i$ are specified for the given experiment.

For the following experiment $r_1 = 1000$, $r_2 = 1000$, $r_3 = 10$. The control was allowed to vary every 10 days. Data was collected every 10 days. The initial state for the problem
was the unhealthy steady state. The initial condition for the estimator was chosen as $x_0^{EKF} = 0.3 \gamma x_0^{true}$ where $\gamma$ is normally distributed with mean 1 and standard deviation 1. Figures 5.14 and 5.15 show the results from this experiment. Note that the EKF is rather effective in filtering the noisy data. This allows for the estimate of the initial condition $x_i$ to the control problem, $P_i$, to be accurate. This ultimately allows the RHC methodology to be successful in transferring the system to the immune response dominant equilibrium.

![Graphs showing control results](image1)

Figure 5.14: Control found when EKF is used as state estimator. Performance of EKF as a filter of noisy data.

### 5.6 State Estimation and Unscheduled Treatment Interruptions

An additional test of this methodology that we would like to consider is the case of using the EKF as a state estimator in addition to interrupting the treatment for 10 days
Figure 5.15: Performance of EKF as a state estimator. Red line denotes EKF estimation, while blue line represents true state.
after 300 days not in accordance with the proposed treatment schedule. Data is collected every 5 days while the control is allowed to vary every 5 days. The parameters for the observation noise for the following experiment are $r_1 = 1000$, $r_2 = 1000$, and $r_3 = 10$. The initial condition for the estimator is chosen as $x_0^{EKF} = 0.3 \gamma x_0^{true}$ where $\gamma$ is normally distributed with mean 1 and standard deviation 1. The plots for these simulations can be seen in Figures 5.16 and 5.17. These results are encouraging. We were able to transfer from unhealthy steady state to healthy in the presence of unscheduled treatment interruption and noisy data collected every five days within approximately 800 days.

Figure 5.16: Control found when EKF is used as state estimator. Performance of EKF as a filter of noisy data. Unscheduled 10 day treatment lapse after 300 days.

### 5.7 Inaccurate Model Parameters

Another natural question to pose especially when considering feedback control is how robust this methodology is with respect to inaccurate models. As this or any other
Figure 5.17: Performance of EKF as a state estimator. Red line denotes EKF estimation, while blue line represents true state. Data was collected every 5 days. Unscheduled 10 day treatment lapse after 300 days.
model is only an approximation of the true physical/biological process, we will study the effectiveness of a dosing strategy based on an inaccurate model. The experimental set up will remain the same as in Section 5.3 except for the fact that our treatment schedule will be based on an inaccurate model. What we mean by inaccurate model is that we will assume the true system is governed by equation (4.1) with model parameters (values in Table 5.1) that vary randomly in each time interval. In computing the control we will not use these randomly perturbed parameters, but rather the values in Table 5.1. The following list will further illustrate this experiment. Let \( q \) denote the vector of model parameters.

1. Using the model, \( \dot{x} = f(x, q_{\text{tab}}, u) \), where \( q_{\text{tab}} \) are the parameter values from the table, solve problem \( P_i \) for optimal control \( u_i \).

2. Create the random set of parameters which will define the true state of the patient for the time interval \([t_i, t_{i+1}]\), using the formula \( (q_{\text{pat}})_j = (\gamma)_j (q_{\text{tab}})_j \) where \((\cdot)_j\) is the \( j^{th} \) component of a vector and each component of \( \gamma \) is normally distributed with mean 1 and a given standard deviation.

3. The model, \( \dot{x} = f(x, q_{\text{pat}}, u) \), is used to simulate the response of the patient on the given time interval.

4. The process is repeated on the following time interval.

For the model responses and controls in Figure 5.18 the standard deviation of \((\gamma)_j\) is 0.05. The results when using inaccurate model parameters were far less promising than those for unscheduled treatment interruptions of when using noisy data. To get the results in Figure 5.18 we had to frequently change doses (daily) and the level of perturbation of the parameters had to be low. Experiments in which more noise was added to the parameters or a less frequent change in dosing was allowed were generally unsuccessful in the transferring the system to healthy steady state.

This behavior should not entirely be unexpected. If we recall the results from the sensitivity analysis in Section 4.6 then we know that this model, especially when the immune response is stimulated, is much more sensitive to variation in parameters than
variation in initial conditions. Thus we should expect our methodology to overcome imperfect state knowledge much more readily than inaccurate parameters.

Figure 5.18: Model response and control when control is allowed to vary daily and model parameters are randomly perturbed.
Chapter 6

Drug-Resistant HIV Model

6.1 Model Description

As an additional test of the RHC methodology we have decided to consider a model similar to the one we have previously studied, based on the modelling efforts of (Bonhoeffer et al. 2000, Callaway & Perelson 2002) but with compartments accounting for drug-resistant strains. We will follow a process similar to one presented in (Kutch & Gurfil 2002, Rong, Feng & Perelson 2007b, Wodarz & Lloyd 2004) amongst others in which an existing model of HIV dynamics is augmented with drug-resistant compartments to create an alternate model of HIV dynamics. The model is as follows

\begin{align*}
\dot{T} &= \lambda - dT - (1 - \epsilon_1)kVT - kV_R T - (1 - \epsilon_1)kV_PT \\
\dot{T}^* &= (1 - \epsilon_1)kVT - \delta T^* - mET^* \\
\dot{V} &= (1 - \epsilon_2)N_T \delta T^* - (c + (1 - \epsilon_1)\rho kT)V \\
\dot{T}_R &= kV_R T - \delta T_R^* - mET_R^* \\
\dot{V}_R &= (1 - \epsilon_2)N_T \delta T_R^* - (c + \rho kT)V_R \\
\dot{T}_P &= (1 - \epsilon_1)kV_PT - \delta T_P^* - mET_P^* \\
\dot{V}_P &= N_T \delta T_P^* - (c + (1 - \epsilon_1)\rho kT)V_P \\
\dot{E} &= \lambda_E + b_E \frac{T^{**}}{T^{**} + K_0} E - d_E \frac{T^{**}}{T^{**} + K_d} E - \delta_E E, 
\end{align*}

(6.1)
where $T$ are the uninfected CD4+ T-cells, $T^*$ are the CD4+ T-cells infected by a drug-sensitive strain of the virus, $V$ is the amount of drug-sensitive HIV, $T^*_R$ are the CD4+ T-cells infected by a strain of HIV which is completely resistant to RTI, $V_R$ is the amount of virus which is immune to RTI, $T^*_P$ are the CD4+ T-cells infected by a strain of HIV which is completely resistant to PI, $V_P$ is the amount of virus which is immune to PI, and $E$ is the body’s own HIV specific immune response. $T^{**} = T^* + T^*_R + T^*_P$ is the total number of infected T-cells.

This model contains several key components of HIV infection and treatment that we will describe. We are primarily interested in the relationship between an HIV specific immune response and drug-resistant strains of the virus. For this reason we have attempted to consider a minimal number of mechanisms that adequately describe this relationship, e.g., we will only consider a single type of target cell and we will not consider any sort of latently infected cell.

Similar to the treatments considered in previous chapters we will consider two types of treatments in this model, RTIs and PIs. In summary, RTIs interfere with the viral RNA-to-DNA synthesis thus blocking the introduction of viral DNA into the cells genetic information. In essence this treatment stops cells from being infected. PIs attack the virus at the stage at which infected cells assemble new viral cells thus producing non-infectious viral cells. In our model we will use the terms $\epsilon_1$ to describe the efficacy of the RTI in stopping infection and $\epsilon_2$ to describe the efficacy of the PI in stopping the creation of new virus by infected cells.

$T$ denotes the uninfected CD4+ T-cells. These cells are created at a rate of $\lambda$ and naturally die proportionally to the number of cells with a rate of $d$. These cells can also be infected by any of the three types of virus at a rate proportional to the product of the number of virus and number of healthy T-cells with a infectivity coefficient of $k$. Note that this rate of infection is reduced by the RTI by the amount $(1 - \epsilon_1)$ for infection by drug-sensitive virus and PI-resistant virus.

$T^*$, $T^*_R$ and $T^*_P$ denote the T-cells infected by drug-sensitive virus, RTI-resistant virus, and PI-resistant virus, respectively. The same term that describes the infection of healthy cells is used to describe the creation of each type of infected cell. Once again note the RTI-resistant strain is not affected by the RTI. These infected cells also die due to viral
infection at a rate proportional to the number of infected cells with death rate of $\delta$. Note that this term will also be used to describe the creation of new free virus. These cells are also destroyed by the HIV specific immune response at a rate proportional to the product of the number of infected cells and the number of the immune effectors at a rate of $m$.

$V, V_R$ and $V_P$ denote the number of free virus infected by the drug sensitive strain, the RTI-resistant strain, and the PI-resistant strain, respectively. These virions are created at rate proportional to the number of infected cell deaths due to viral infection with a rate of $N_T$ free virus produced per cell. Note that in the drug sensitive case and the RTI-resistant case the creation of free virus is reduced by the PI at a rate of $(1 - \epsilon_2)$. The free virions die at a rate proportional to the number of free virus at a rate of $c$ per day. The free virions also are eliminated by infection of a healthy cell. The term $-\rho k TV$ accounts for this. We choose $\rho = 1$ to describe a single virion being responsible for an infection.

$E$ denotes the HIV specific immune response. Rather than model immune memory we will use a source term $\lambda_E$ to create an non-zero off-treatment steady state for $E$. These cells also naturally die at a rate proportional to the number of immune effectors at a rate of $\delta_E$. We will model the HIV related birth and death of these cells with the terms $b_E F(T^{**}) E$ and $d_E G(T^{**}) E$, respectively. $F$ and $G$ are Michaelis-Menten type saturation terms. Simply stated this sort of modelling of the HIV specific immune response suggests that this response is stimulated in proportion to the number of immune effectors and infected T-cells when the number of infected T-cells is not very large or small, however when there are a large or small number of infected T-cells this response is diminished.

Table 6.1 contains the values for the model parameters used in all simulations, which are based on values presented in (Bonhoeffer et al. 2000, Callaway & Perelson 2002). They were taken from the previous model, with the target cells in this model corresponding to the first kind of target cells in the previous model.
6.2 Model Discussion

As earlier stated, the primary purpose of this model is to study the relationship between the HIV specific immune response and drug-resistant viral strains. This relationship has been studied in (Shiri, Garira & Musekwa 2005, Wodarz & Lloyd 2004); however, the mechanism by which the immune response is stimulated is different in these works. This stimulation is modelled by the terms $TVE$ and $T^*E$ respectively as opposed to the Michaelis-Menten saturation term we have used. As such we have proposed a model that focuses primarily on this mechanism and have made simplifications in other areas of modelling the in vivo dynamics of HIV. We would be remiss if we did not discuss some of the assumptions implicit in this model.

1. This model only accounts for one type of target cell. Target cells could be differentiated according to location or cell type, each with potentially different model parameters describing susceptibility to infection, efficacy of drug, etc.

2. This model does not account for acutely infected cells or any sort of delay between infection and production of new virus. It assumes that as soon as a cell is infected it immediately becomes active in the production of new virus.

3. It is assumed that the only difference in the various strains of virus is their immunity to certain forms of treatment. Thus, many of the parameters related to infection rates, death rates, and such (e.g., $k$, $\delta$) are the same for all infected cell types and
all virus types. There is biological evidence that suggests that resistant strains of virus are less fit than the drug sensitive strains (Archer et al. 2000, Hance et al. 2001). Generally speaking, if this were not the case the resistant case would out compete the wild-type even in the absence of treatment and become the dominant strain. As such, models presented in (Snedecor 2003, Rong et al. 2007b) model the resistant strains as less infective and producing fewer number of virions on average with reduced values of $k$ and $N_T$ for the resistant strains.

4. This model does not include a specific compartment for non-infectious virus as this compartment does not affect other compartments and is mainly useful in designing estimators and working with measurements. As we will not be using these techniques in this chapter we will omit this compartment.

5. There is no mechanism in this model for drug-sensitive virus to become drug-resistant. This can be modelled deterministically as a proportion of the number of infections, i.e., this rate would be proportional to $VT$, e.g., see (Rong et al. 2007b).

6. This model also assumes drug resistance is absolute, i.e., a strain of virus is either completely immune to a form of treatment or is completely susceptible. This is perhaps an overestimation of the resistance. This is often modelled as the efficacy of the drug to the resistant strain $\epsilon_r$, being some proportion of the efficacy of the drug to the sensitive strain $\epsilon_s$ (Snedecor 2003). In models of HIV dynamics that include a pharmacokinetic model of the drug action, e.g. (Dixit, Markowitz, Ho & Perelson 2004, Rong et al. 2007b), this decrease in efficacy is modelled as an increase in the intracellular concentration of the drug needed to inhibit viral replication by 50%.

7. We have assumed that no strain is immune to both RTIs and PIs.

In spite of these simplifications, this model does offer the potential to study the feasibility of using optimal control methodology, specifically a RHC approach, to study the stimulation of the HIV specific immune response in the presence of drug-resistant viral strains. Our model considers completely resistant strains of the virus to both general
classes of drugs that are as fit as the wild-type virus. If this control methodology can be shown effective in a model with these characteristics then this would demonstrate the potential of this methodology in a more realistic model.

6.2.1 Model analysis in the absence of treatment

This model also simplifies in the absence of drug treatment. If we assume \( \epsilon_i = 0, \ i = 1, 2 \), and if we write \( V^{**} = V + V_R + V_P \) and \( T^{**} = T^* + T_R^* + T_P^* \) we can write our model as

\[
\begin{align*}
\dot{T} &= \lambda - dT - kV^{**}T \\
\dot{T}^{**} &= kV^{**}T - \delta T^{**} - mET^{**} \\
\dot{V}^{**} &= N_T \delta T^{**} - (c + \rho kT)V^{**} \\
\dot{E} &= \lambda_E + b_ET^{**}T^{**} + K_E - d_ET^{**}T^{**} + K_dE - \delta_EE.
\end{align*}
\]  

(6.2)

As we would like to study whether HIV infection can be controlled by the body’s own virus specific immune response, in the absence of medication, we will present some analysis of this reduced model. In particular, we would like to study equilibrium solutions and their stability properties. We found three physically realistic equilibria. There is the obvious equilibrium solution

\[
T^* = \frac{\lambda}{d} \text{ cells/mm}^3, \ T^{**} = 0 \text{ cells/mm}^3, \ V^{**} = 0 \text{ virions/mm}^3, \ E^* = \frac{\lambda_E}{\delta_E} \text{ cells/mm}^3,
\]  

(6.3)

which corresponds to an uninfected individual. The Jacobian of the equation was computed at the equilibrium using AD. The Jacobian has an eigenvalue with positive real parts and thus this equilibrium is unstable. This can be seen in simulations starting close to this equilibrium.

Two other steady states were found numerically. They were found approximately by simulating the model with various initial conditions. Once approximate equilibria were found, then Newton’s method with forward difference derivatives and direct factorization of the linear equations was used to find the true equilibria. The code used
for in this problem was written by Tim Kelley and can be obtained at his home page at http://www.math.ncsu.edu/. For more information on solving nonlinear equations in this manner see (Kelley 2003). The first equilibria is

\[ T = 164.1971 \text{ cells/mm}^3, \ T^* = 11.9360 \text{ cells/mm}^3, \]
\[ V^* = 63.6280 \text{ virions/mm}^3, \ E = 0.0236 \text{ cells/mm}^3, \] (6.4)

which we will term the unhealthy equilibrium. Note that there is a small amount of healthy T-cells, a large amount of infected cells and free virus, and the immune response is limited. The Jacobian evaluated at this equilibrium has negative real parts of all eigenvalues and thus this equilibrium is stable.

Another equilibrium is

\[ T = 967.6984 \text{ cells/mm}^3, \ T^* = 0.0821 \text{ cells/mm}^3, \]
\[ V^* = 0.4172 \text{ virions/mm}^3, \ E = 323.4259 \text{ cells/mm}^3, \] (6.5)

which we will term the healthy equilibrium. Note that there is a large amount of healthy T-cells, minimal infected cells and free virus and the HIV specific immune response is strong. The Jacobian evaluated at this equilibrium has negative real parts of all eigenvalues and thus this equilibrium is stable.

Finally, we will note that as the dynamics of the model are fundamentally similar to the previously considered model of HIV dynamics one can prove the existence and uniqueness of solutions to this model and the fact that solutions to this model with non-negative initial data have are non-negative using the same methods and arguments given in Section 4.3. Recall that this required the introduction of an alternate model that accounted for saturated dynamics. As before this saturated model was not used in numerical simulations.

6.3 Model Simulations

Simulations with this model illustrate reasonable behavior. If there is no treatment then the dynamics are similar to the model this was derived from. If only one type of
treatment is used then the virus resistant to this type of treatment dominates. Figure 6.1 shows the case in which only PI is administered and Figure 6.2 shows the case in which only RTI is administered. Note that when only PIs are administered the PI-resistant virus dominates and similarly when only RTIs are administered the RTI-resistant virus dominates. In both cases the drug-sensitive virus is largely suppressed. This is most likely due to the fact that the drug-sensitive strain has no “fitness” advantage. As a note the initial condition for the model is similar to an acute infection that we previously considered in that only 1 percent of the virus was considered to be RTI-resistant and PI-resistant. Numerically, this initial condition is

\[
T = 1000 \text{ cells/mm}^3, \quad T^* = 0 \text{ cell/mm}^3,
\]

\[
T_R^* = 0 \text{ cells/mm}^3, \quad T_P^* = 0 \text{ cell/mm}^3,
\]

\[
V = 0.001 \text{ virions/mm}^3, \quad V_P = 10^{-5} \text{ virions/mm}^3, \quad V_R = 10^{-5} \text{ virions/mm}^3,
\]

\[
E = 0.024 \text{ cells/mm}^3.
\]

### 6.4 Optimal Control and Necessary Conditions

In our attempt to stimulate the immune response, ideas from optimal control will be employed. We will mimic the approach from the drug sensitive model and attempt to find a treatment regimen that minimizes the objective function

\[
J(\epsilon_1, \epsilon_2) = \int_{t_0}^{T_f} \left[ q_1 V + q_2 V_R + q_3 V_P + r_1 \epsilon_1^2 + r_2 \epsilon_2^2 - SE \right] dt + \alpha(E(T_f) - 323.4259)^2. \tag{6.7}
\]

Note that \( \epsilon_1 \) is bounded between 0 and \( \epsilon_{1,\max} \) and \( \epsilon_2 \) is bound between 0 and \( \epsilon_{2,\max} \). Note that our cost seeks to minimize the total free virus, the control action used, and the difference between the final immune response and the healthy steady state immune response while maximizing the immune response. Also note that if we write

\[
x = [T \quad T^* \quad V \quad T_R^* \quad V_R \quad T_P^* \quad V_P \quad E]^T
\]

and

\[
u = [\epsilon_1 \quad \epsilon_2]^T
\]

then the cost can be written

\[
J(u) = \int_{t_0}^{T_f} L(x, u, t) dt + \phi(x(T)). \tag{6.8}
\]
Figure 6.1: Only PIs administered.
Figure 6.2: Only RTIs administered.
The existence and uniqueness to this optimal control problem can be shown by a similar argument as the result in Section 4.4 and the added condition that $\phi$ is continuous. As $\phi$ is quadratic in $E$ this condition holds. The Hamiltonian and necessary conditions are as given in equations (4.20), (4.21) and (4.22).

For completion we will include these necessary conditions here. The Hamiltonian is

$$H = q_1 V + q_2 V_R + q_3 V_P - SE + r_1 \epsilon_1^2 + r_2 \epsilon_2^2$$

$$+ p_1(\lambda - dT - (1 - \epsilon_1)kVT - kV_R T - (1 - \epsilon_1)kV_P T)$$

$$+ p_2((1 - \epsilon_1)kVT - \delta T^* - mET^*)$$

$$+ p_3((1 - \epsilon_2)N_T \delta T^* - (c + (1 - \epsilon_1)\rho kT)V)$$

$$+ p_4(kV_R T - \delta T^*_R - mET^*_R)$$

$$+ p_5((1 - \epsilon_2)N_T \delta T^*_R - (c + \rho kT)V_R)$$

$$+ p_6((1 - \epsilon_1)kV_P T - \delta T^*_P - mET^*_P)$$

$$+ p_7(N_T \delta T^*_P - (c + (1 - \epsilon_1)\rho kT)V_P)$$

$$+ p_8(\lambda E + b_E \frac{T^{**}}{T^{**} + K_b} E - d_E \frac{T^{**}}{T^{**} + K_d} E - \delta_E E).$$

Now we will differentiate the Hamiltonian with respect to the state variables to obtain
the equations for the costate variables

\[
\begin{align*}
\dot{p}_1 &= -p_1(-d - (1 - \epsilon_1)kV - kV_R - (1 - \epsilon_1)kV_P) \\
&\quad - p_2(1 - \epsilon_1)kV + p_3(1 - \epsilon_1)\rho kV - p_4kV_R + p_5\rho kV_R \\
&\quad - p_6(1 - \epsilon_1)kV_P + p_7(1 - \epsilon_1)\rho kV_P \\
\dot{p}_2 &= -p_2(-\delta - mE) - p_3(1 - \epsilon_2)N_T \delta \\
&\quad - p_8 \left( \frac{b_E}{T^* + T_R^* + T_P^* + K_b} - \frac{b_E(T^* + T_R^* + T_P^*)}{(T^* + T_R^* + T_P^* + K_b)^2} \right) \\
&\quad - p_8 \left( \frac{d_E}{T^* + T_R^* + T_P^* + K_d} + \frac{d_E(T^* + T_R^* + T_P^*)}{(T^* + T_R^* + T_P^* + K_d)^2} \right) \\
\dot{p}_3 &= -q_1 + p_1(1 - \epsilon_1)kT - p_2(1 - \epsilon_1)kT - p_3(-c - (1 - \epsilon_1)\rho kT) \\
\dot{p}_4 &= -p_4(-\delta - mE) - p_5(1 - \epsilon_2)N_T \delta \\
&\quad - p_8 \left( \frac{b_E}{T^* + T_R^* + T_P^* + K_b} - \frac{b_E(T^* + T_R^* + T_P^*)}{(T^* + T_R^* + T_P^* + K_b)^2} \right) \\
&\quad - p_8 \left( \frac{d_E}{T^* + T_R^* + T_P^* + K_d} + \frac{d_E(T^* + T_R^* + T_P^*)}{(T^* + T_R^* + T_P^* + K_d)^2} \right) \\
\dot{p}_5 &= -q_2 + p_1kT - p_4kT - p_5(-c - \rho kT) \\
\dot{p}_6 &= -p_6(-\delta - mE) - p_7N_T \delta \\
&\quad - p_8 \left( \frac{b_E}{T^* + T_R^* + T_P^* + K_b} - \frac{b_E(T^* + T_R^* + T_P^*)}{(T^* + T_R^* + T_P^* + K_b)^2} \right) \\
&\quad - p_8 \left( \frac{d_E}{T^* + T_R^* + T_P^* + K_d} + \frac{d_E(T^* + T_R^* + T_P^*)}{(T^* + T_R^* + T_P^* + K_d)^2} \right) \\
\dot{p}_7 &= -q_3 + p_1(1 - \epsilon_1)kT - p_6(1 - \epsilon_1)kT - p_7(-c - (1 - \epsilon_1)\rho kT) \\
\dot{p}_8 &= S + p_2\alpha T^* + p_3\alpha T_R^* + p_4\alpha T_P^* \\
&\quad - p_8 \left( \frac{b_E(T^* + T_R^* + T_P^*)}{(T^* + T_R^* + T_P^* + K_b)} - \frac{d_E(T^* + T_R^* + T_P^*)}{(T^* + T_R^* + T_P^* + K_d)} \right) - \delta E.
\end{align*}
\]

with terminal conditions \( p_i(T_f) = 0 \) for \( i = 1, ..., 7 \) and \( p_8(T_f) = 2\alpha(E(T_f) - 323.4259) \).

In addition, we will also need to compute the derivative of the Hamiltonian with
respect to the control variables for use in our numerical optimization routines

\[
\frac{\partial H}{\partial \epsilon_1} = 2r_1 \epsilon_1 + p_1 (kVT + kV_PT) - p_2 kVT + p_3 \rho kTV - p_6 kV_PT + p_7 \rho kT \]

\[(6.11)\]

\[
\frac{\partial H}{\partial \epsilon_2} = 2r_2 \epsilon_2 - p_3 N_T \delta T^* - p_5 N_T \delta T^R. \]

\[(6.12)\]

Note that these were computed symbolically with Maple. AD was not used as the iterative
method of computing the optimal control is computationally expensive and the increase
in computational speed afforded by the analytic calculation of these derivatives was
considered worth the effort.

6.5 RHC

As the development of this model was based around a question of whether the results
from the earlier work were possible to repeat when we had to consider the possibility
of drug resistance we will present these results. The RHC methodology is the same as
that presented in the drug-sensitive case, Section 5.2, with parameters given in Table 6.2.
The initial conditions for this simulation are very similar to what we have termed the
unhealthy equilibrium with non-zero levels of both strains of drug-resistant virus. The
specific values for this initial condition are as follows

\[
T = 164.2 \text{ cells/mm}^3, \quad T^* = 9.93 \text{ cell/mm}^3,
\]

\[
T^*_R = 1 \text{ cells/mm}^3, \quad T^*_P = 1 \text{ cell/mm}^3,
\]

\[
V = 61.6 \text{ virions/mm}^3, \quad V_P = 1 \text{ virions/mm}^3, \quad V_R = 1 \text{ virions/mm}^3
\]

\[
E = 0.024 \text{ cells/mm}^3.
\]

(6.13)

The results can be seen in Figures 6.3 and 6.4. Note that in these simulations both
the RTI and PI efficacy were bounded above by 0.8. This control represents a unique
version of STI. Note that the PI control cycles on and off 3 times in the first 200 days.
Note the increase in RTI-resistant virus during these periods. Then between 200 and
400 days the methodology calls for approximately 3 interruptions of the RTI while the
Table 6.2: Parameters for RHC problem.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_{i}$</td>
<td>50 $i$ days</td>
</tr>
<tr>
<td>$t_{ch,i}$</td>
<td>50 days</td>
</tr>
<tr>
<td>$q_1$</td>
<td>$5 \times 10^{-4}$</td>
</tr>
<tr>
<td>$q_2$</td>
<td>$5 \times 10^{-4}$</td>
</tr>
<tr>
<td>$q_3$</td>
<td>$5 \times 10^{-4}$</td>
</tr>
<tr>
<td>$r_1$</td>
<td>$5 \times 10^{-4}$</td>
</tr>
<tr>
<td>$r_2$</td>
<td>$5 \times 10^{-4}$</td>
</tr>
<tr>
<td>$S$</td>
<td>$10^{-1}$</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>$5 \times 10^{-3}$</td>
</tr>
</tbody>
</table>

PI is kept at a moderately level of efficacy. The PI-resistant virus rebounds during this period. Note that the drug-sensitive strain is generally kept at low levels by one drug or the other. This dynamic could potentially be more complicated or different if one were to consider a fitness cost for the viral resistance.

Now as an attempt to reduce the amount of efficacy used and increase the amount of time before the control can be changed and still transfer the system to the healthy steady state we performed a number of simulations. We will present the results for when we allowed the control to vary only every 5 days and the RTI efficacy was bounded by 0.7 and the PI efficacy was bounded by 0.6. These results can be seen in Figures 6.5 and 6.6. Note that the RTI is kept at a relatively high level while the PI is cycled on and off.

### 6.6 Unscheduled Treatment Interruption

We would also like to consider the situation in which treatments are interrupted in contradiction to the designed treatment regiment. We will consider a situation very similar to that presented in Section 5.4. For the following simulation the RTI and PI efficacy bound are both 0.8 and treatment is allowed to vary daily. Treatment is stopped for 10 days after 200 days. The results for these simulations can be seen in Figures 6.7 and 6.8. As a chief indicator of the effect the interruption had we will consider the response of the immune effector in this case and the case in which the regiment was followed. See
Figure 6.3: Model response for drug-resistant model when control is allowed to vary daily.
Figure 6.4: Control for drug-resistant model when the control is allowed to vary daily.
Figure 6.5: Model response for drug-resistant model when control is allowed to vary every 5 days.
Figure 6.6: Control for drug-resistant model when control is allowed to vary every 5 days.
Figure [6.9] for this comparison. Note that this interruption increased the time it took to reach healthy steady state from approximately 500 days to approximately 900 days. This is a larger increase than we saw in the drug-sensitive case.

Figure 6.7: Model response for drug-resistant model when control is allowed to vary daily and control was interrupted after 200 days for 10 days.
Figure 6.8: Control for drug-resistant model when control is allowed to vary daily and control was interrupted after 200 days for 10 days.
Figure 6.9: Comparison of immune effector response when control was interrupted and when it was followed.
Chapter 7

Summary and Future Work

7.1 Summary

The 2D research in electron gun design demonstrated that computer optimization could be effectively applied to electron gun design. Using the procedures established in this research, it should be possible for a design engineer to set up an optimization process achieving specified electron gun voltage, current, beam size, and beam ripple goals in approximately two hours. Assuming, for example, that the engineering rate is $150 per hour, this would cost approximately $300. Compare this with the engineering time required to perform these tasks manually. For typical engineering time of 30 hours, the cost would be $4,500. Assuming a similar percentage gain for the sheet beam gun developed by CCR, computer optimization would result in a savings of $56,000.

If 3D optimization techniques can be successfully developed and if our initial work with the SBG is any indication then, it will enable a new generation of devices that can not currently be designed. This will allow new innovative applications for electron beam devices.

In our work on the optimal design of TWTs we described additional tools to effectively explore different circuit designs, automate parameter variation, and optimization. In particular, we demonstrated that accurate simulation codes, such as the physics-based large-signal code CHRISTINE-1D, can be seamlessly integrated with state-of-the-art optimization codes, such as DIRECT and/or Nelder-Mead algorithms. By treating simulation and optimization codes as "black boxes", one can explore different circuit designs using combinations of optimization criteria formulated through the goal (objective) functions. We demonstrated that our proposed optimization tool, the Nelder-Mead algorithm, appears to be more robust when compared to the CHRISTINE-1D built-in efficiency optimization algorithm for this circuit design. This may be particularly useful in the optimization of TWT multistage depressed collector stages where a good initial iterate is not always available. In addition, we also applied our optimization tools to the design of a folded waveguide slow-wave circuit. We demonstrated that CHRISTINE-1D with DIRECT and Nelder-Mead in a hybrid optimization approach can improve the folded waveguide slow-wave circuit performance based on several frequency dependent optimization criteria. These criteria include saturated RF efficiency, small signal gain and phase difference minimization, and desired output power at saturation.

In our work with HIV we have discussed a model that exhibits many fundamental properties of HIV infection, including multiple steady states, compartments for multiple target cells, and multiple methods of virus control such as RTIs, PIs, and CTL response. We illustrated how numerical optimal control techniques can be used to develop treatment schedules which suggests an STI type dosing strategy. We then analyzed the sensitivity of the model parameters under this treatment schedule, an exercise that suggests the bodies own immune response is very important under this type of treatment schedule. We finally described how stochastic estimation techniques can be used to both estimate unknown states and parameters.

We then introduced RHC methodology and stochastic estimation, in particular the
EKF. It has been shown that these techniques can be used to derive treatment strategies that can effectively control an HIV infection for our present model. These techniques have been shown to be effective with respect to how often the drug dosing can vary, whether the patient effectively adheres to the treatment strategy and data corrupted by noise.

Finally, we introduced a model of HIV dynamics that include drug-resistant strains. As with the previous HIV model we have been able to use RHC methods to transfer the system to an immune dominant steady state. We have also shown that this methodology is able to overcome poor drug adherence.

7.2 Future Work in the Design of Electronic Devices

The ultimate goal of this research is to develop design tools for complex, 3D, electron beam devices that provide improved capabilities or performance over standard 2D devices. In the microwave tube industry, significant research is under way to utilize distributed beam devices, including sheet beam and multiple beam configurations. Distributed beams allow significant reduction in operating voltage, which translates to dramatically lower cost power supplies and improved operating efficiency. It is also anticipated that a new generation of electron devices could be developed for new and innovative applications.

The work in designing electronic devices has shown us a variety of areas for research that can largely be grouped into two categories: the further refinement of the BOA software, especially as it relates to optimization and the use of this software to design more complex electron devices.

In regards to the first category there are several issues to address. We would like to add to BOA the capability of creating magnetic fields, rather than importing them from other programs. With this capability it would be simpler to manipulate this aspect of the design and thus one could use magnetic field strength and location as a design parameter. In addition it would be possible to consider a magnetic field that varied with axial location. We would also like to consider the economic aspects of specific device design. The ability to apply optimization to create designs with reduced fabrication costs
has obvious financial benefits. We would also like to integrate the optimization routines described in this work into the BOA code. In the current state our design routines require a good deal of user knowledge, both related to the optimization routines (DIRECT, Nelder-Mead, implicit filtering) and software (BOA, Solidworks®, MATLAB®). If these could be integrated into the BOA software it could be a powerful design tool. Implicit filtering is readily implemented in parallel. This could significantly increase the speed of the design process. Finally, the model simulations in question seem to be a good place to consider surrogate model optimization. One could initial reduce the convergence criteria and coarsen the mesh in the attempt to quickly approximate a good design. Then the tolerances could be increased to create a realistic design. A theoretically sound and readily implementable methodology for using this technique would be a significant contribution to this area. Ultimately the goal of this part of the development of BOA is the integration of all the above areas to create generalized optimization routines for electron device design.

In 2005, the U.S. Department of Energy awarded CCR a contract to develop a doubly convergent MBG for an L-Band klystron. Figure 7.1 illustrates a potential design for a MBG. Note there are eight emitters on the left side of the figure each with individual beam tunnels closer to the central axis of the device. The current generation of multiple beam electron guns use singly convergent beams. This means the beams converge about their individual axes, defined by the cathodes, but do not converge toward the axis of the device. In the singly convergent guns the emitters are not at larger radius from the central axis of the device than the beam tunnels. Design of MBGs that converge about the beams’ local axis and the device axis are referred to here as doubly convergent MBGs. After several months of development, it became apparent that the number of variables exceeded capabilities to manually design such a device. Consequently, the effort was abandoned, and CCR initiated the current program to develop computer optimization techniques for such 3D devices. As the individual beams converge about their local axis, they begin to rotate about that axis to create the $v \times B$ force necessary to balance space charge forces within the beam. Nevertheless, the beams continue to propagate parallel to the device axis in singly convergent guns. For doubly convergent beams, the compression toward the device axis causes the beams to rotate about the axis of the device, following a
spiral path. This spiraling about the device axis makes design and fabrication of complex circuits impractical, at least using currently available fabrication techniques.

Figure 7.1: Example of potential design for MBG.

The rotation of doubly convergent MBG beams about the device axis is necessary to conserve angular momentum. If the electrons could be emitted from the cathodes at an angle to the device axis (and magnetic field), they could possibly satisfy conservation requirements with the beams propagating parallel to the device axis when fully compressed. Angular injection at the cathode will require complex, 3D structures to preserve the beam quality. Following completion of the sheet beam gun effort, CCR will use optimization techniques to design these complex structures.

Successful development of doubly convergent MBGs would allow design of high power,
multiple beam klystrons using fundamental mode circuits. Existing multiple beam, high power devices must use overmoded circuits to avoid excessive emission current densities at the cathode. Not only does this complicate the design, but increases the radial size and potential for parasitic mode generation. The increased size of the device also increases the size of the magnetic circuit and, hence, its power supply requirements.

As an initial step in the design of MBGs a project, which I am participating in, is being carried out at The NC State Research Experiences for Undergraduates in mathematics: Modeling and Industrial Applied Mathematics.

7.3 Future Work in Relation to HIV

Several questions have been raised in regard to our work with HIV. First, our sensitivity analysis is only local, i.e., our simulations are for a specific set of model parameters and control strategy. We would like to understand how do these sensitivities vary as a function of both parameter and control. The local results already include 189 relative and generalized sensitivities, each of which are functions of time. A manageable way to understand these results could offer insights into the model dynamics and treatment potential.

Our control work considers drug action in terms of "efficacy". The use of a pharmacokinetic model for specific drug dynamics would be more realistic and could bring this methodology closer to being a clinical tool for physicians. Our drug-resistant model does not account for the development of resistance or varying levels of resistance. If one were to account for these dynamics the predicted control will change. An important issue to consider would be as to whether the underlying form of some sort of STI would remain.

In regards to our estimation problem, the data was entirely simulated. An obvious next step would be to incorporate real data. However, real data for these models is often censored, i.e, the tests cannot detect a viral load below a certain threshold. The current implementation of our filter has no mechanism to account for this other than simply ignoring the data. Recently, a filter, which accounts for this issue, has been developed (Ibarz-Gabardos & Zufiria 2005). This could improve the performance of these filters on real data.
Our model captures several key mechanisms of HIV infection, but obviously omits others. A promising new idea is the use of an age-structure model (Rong, Feng & Perelson 2007a, Nelson, Gilchrist, Coombs, Hyman & Perelson 2004). This allows for the death and viral production of infected cells to depend on the age of the infected cell. This is a generalization of the latent infections and delay terms used in many HIV models. The drawback of this methodology is that the model equations are now PDEs instead of ODEs and the death and cell production rates are functions which must be estimated instead of constant parameters.
Bibliography


*CST Microwave Studio Suite* (2005).


*U. S. Patent 2,3000,052.* (1942).


APPENDICES
Appendix A

Optimization Methods

As a large portion of this work deals with the implementation of various optimization routines, we have included a description of these methods as an appendix.

A.1 Nelder-Mead Algorithm

The Nelder-Mead algorithm is a deterministic sampling method, i.e., it only requires function values and no gradient information and is thus a simple algorithm to implement. Given an $N+1$ simplex of points in $\mathbb{R}^N, \{\lambda_1, \lambda_2, \cdots, \lambda_{N+1}\}$ it first sorts the simplex points such that $J(\lambda_1) \leq J(\lambda_2) \leq \cdots \leq J(\lambda_{N+1})$. It then attempts to minimize the function by replacing the simplex point with highest function value, the worst point $\lambda_{N+1}$, with a point with a lower function value. It first computes the centroid of the simplex, not including the worst points

$$\bar{\lambda} = \frac{1}{N} \sum_{i}^{N} \lambda_i. \quad (A.1)$$

It then attempts to replace $\lambda^{N+1}$ with

$$\lambda_{NEW} = (1 + \beta)\bar{\lambda} - \beta \lambda_{N+1}, \quad (A.2)$$

where

$$\beta = \{\beta_r, \beta_e, \beta_{oc}, \beta_{ic}\} \quad (A.3)$$
and the values for $\beta$ taken from
\[-1 < \beta_{ic} < 0 < \beta_{oc} < \beta_r < \beta_e.\]

If none of these values are better than the previous worst point, the algorithm shrinks
the simplex towards the best point, i.e., it replaces the points with
\[\lambda_i = \frac{\lambda_i + \lambda_1}{2}.\]

The algorithm then resorts the points and iterates. Our implementation of the algorithm
used the values
\[\{\beta_r, \beta_e, \beta_{oc}, \beta_{ic}\} = \{1, 2, 1/2, -1/2\}. \quad (A.4)\]

To better illustrate this method, consider the case when $N = 2$. In this case the parameter
space is $\mathbb{R}^2$ and our simplex is a triangle in the plane. Figure [A.1] illustrates what Nelder-
Mead does in this case. Note $\lambda_3$ is the worst parameter and the algorithm is attempting
to replace this point with either a reflection, $r$, expansion, $e$, outward contraction, $oc$, or
inward contraction, $ic$. There are various stopping criteria for this algorithm including
iteration number, difference between the function at the best and worst points, and
simplex size. For a more detailed treatment see (Kelley 1999).

### A.2 Implicit Filtering

Implicit filtering is a projected quasi-Newton iteration that uses difference gradients,
reducing the difference increments as the optimization progresses (Kelley 2002). Note
that BFGS is an example of a quasi-Newton method which we will describe in Section
A.4. The idea is that gradient approximations with a large step size will be insensitive
to high-frequency oscillations, which generally produce a large number of local minima,
and follow the general landscape of the parameter space. As the routine approaches
the minimum, where the oscillations in the parameters space are less, the step size for
the gradient is reduced. Figure [A.2] illustrates a function where implicit filtering may
be useful. The specific implementation used in this work is bound constrained, but, for
simplicity, the unconstrained version is described. This algorithm optimizes functions of
the form
Figure A.1: Nelder-Mead in 2 dimensions.

\[ f = f_s + \phi, \quad (A.5) \]

where \( f \) is smooth and \( \phi \) is a low-amplitude perturbation that is smallest near the minimum.

Given a current iterate \( \lambda \) and step size \( h \), the algorithm proceeds as follows. First \( f \) is sampled at the \( 2N \) points in the stencil

\[ S(\lambda, h) = \{ \lambda \pm he_i \}, \quad (A.6) \]

where \( e_i \) is the unit vector in the \( i \)th coordinate direction. Let

\[ f^* = \min\{ f(z) | z \in S(\lambda, h) \}. \quad (A.7) \]
If $f^* > f(\lambda)$ then this is called stencil failure and, $h$ is reduced and the routine iterates. If $f^* < f(\lambda)$ then the routine attempts to update $\lambda$ with the formula

$$
\lambda_{NEW} = \lambda - \nu H^{-1} \nabla_h f(\lambda),
$$

where $H$ is the model Hessian obtained from the quasi-Newton update, $\nabla_h f(\lambda)$ is the approximate gradient obtained from using a centered difference formula on the stencil, and $\nu$ is a line search parameter to insure sufficient decrease. This proceeds until $h$ reaches a user specified value or another another convergence criteria such as small norm of simplex gradient or maximum number of function evaluations.

### A.3 DIRECT

The third algorithm we used is DIRECT, which was first introduced in (Jones, Perttunen & Stuckman 1993). DIRECT was designed for simple, bounded, constrained, non-
smooth problems, and, as the optimization progresses, it performs a global search. In the limit, DIRECT will sample a dense grid of points in the feasible set. The name DIRECT stands for DIviding RECTangles, which is very indicative of how the algorithm progresses towards finding the optimum. More specifically, we consider a bounded-constrained optimization problem of the form \( \min_{\lambda \in \Omega} f(\lambda) \), where \( f : \mathbb{R}^N \to \mathbb{R} \) and is Lipschitz continuous on \( \Omega = \{ \lambda \in \mathbb{R}^N \mid L_i \leq \lambda_i \leq U_i \} \subset \mathbb{R}^N \). Since \( \Omega \) has hard bounds, we refer to \( \Omega \) as a hyper-rectangle in \( \mathbb{R}^N \). DIRECT initiates its search by sampling the objective function at the center of \( \Omega \), \( c \). At each iteration, DIRECT method balances the global and local searches by identifying potential optimal hyper-rectangles based on the value of the objective function at the center and the size of the hyper-rectangle. For example, at the first iteration when the hyper-rectangle \( \Omega \) is potentially optimal, all coordinate directions of the hyper-rectangle are long, and DIRECT divides every direction (see Figure A.3). DIRECT employs a simple heuristic argument to determine the order in which long sides (or directions) of the hyper-rectangles are divided. In particular, it evaluates the objective function at \( c \pm \delta e_i / 3, \omega_i = f(c \pm \delta e_i / 3) \), where \( \delta \) is the maximal side length of the hyper-rectangle. It then divides the hyper-rectangle into smaller hyper-rectangles (thirds), starting with the dimension with smallest \( \omega_i \) and continuing to the dimension with the largest \( \omega_i \). In subsequent iterations, it determines which of the hyper-rectangles is potentially optimal and divides them in a similar manner. Once again, to better illustrate this method, we consider the case where \( N = 2 \). Here, hyper-rectangles are merely rectangles, and the division process is easy to visualize. Figure A.3 illustrates this process. For a more detailed treatment of the DIRECT method, we refer the interested reader to (Finkel 2003, Finkel & Kelley 2006).

A.4 BFGS

The BFGS (Broyden, Fletcher, Goldfarb, Shanno) method is based on the idea of forming a quadratic model of the function at your given iterate \( \lambda_c \);

\[
m_c(\lambda) = J(\lambda_c) + \nabla J(\lambda_c)^T (\lambda - \lambda_c) + \frac{1}{2} (\lambda - \lambda_c)^T \nabla^2 J(\lambda_c) (\lambda - \lambda_c).
\]  (A.9)
We then attempt to find a critical value of
\[
\nabla m_c(\lambda) = \nabla J(\lambda_c) + \nabla^2 J(\lambda_c)(\lambda - \lambda_c) = 0,
\]
and if the Hessian of $J$ is positive definite then the critical value is a minimum. Substituting $\lambda_+$ into $\nabla m_c(\lambda) = \nabla J(\lambda_c) + \nabla^2 J(\lambda_c)(\lambda - \lambda_c) = 0$ and solving for $\lambda_+$ gives us the formula for $\lambda_+$. BFGS is a quasi-Newton method in that instead of using the true Hessian matrix at a given iterate, which may or may not be positive definite, it uses an approximation of the Hessian and updates it at each step with a rank-two update. An iterate for BFGS includes the steps
\[
\lambda_+ = \lambda_c + \lambda H_c^{-1} \nabla J(\lambda_c).
\]
with line search parameter $\lambda$. The Hessian is updated with
\[
s = \lambda_+ - \lambda_c
\]
\[
y = \nabla J(\lambda_+) - \nabla J(\lambda_c)
\]
\[ H_+ = H_c + \frac{yy^T}{y^T s} - \frac{(H_c s)(H_c s)^T}{s^T H_c s}. \]  
(A.14)

This method is locally superlinearly convergent if the initial iterate and the initial Hessian are close enough to the minimizer and the Hessian at the minimizer, respectively.
Appendix B

Automatic Differentiation

The computation of derivatives is one of the fundamental exercises in mathematics. However, the accuracy in the computation of the derivative of a high dimensional function can be difficult. If one has a function $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ then the analytic computation of the Jacobian requires $nm$ calculations each of which may be non-trivial. This is tedious error-prone work. A difference approximation may be employed; however, this can be difficult if the components of the function are differently scaled. Another issues with using a difference approximation is that the result is only an approximation and may require at least $n + 1$ function evaluations depending on the scheme used. Symbolic computations can be used; however, the $nm$ expressions can be complicated. (Griewank 2002).

AD offers an alternative to these methods. It is accurate to machine precision and is easy to use. It is a techniques that exploits the chain rule. No matter how complicated the function is, at the most basic level it is composition of elementary functions, such as $+, -, \text{ or } \sin(\cdot)$. The derivatives of these elementary functions are known and the total derivative can be calculated using the chain rule (Verma 2000). In this research, we have used AD to greatly reduce the human labor required to employ some of the techniques presented in this research; such as sensitivity analysis, equilibrium analysis, and filtering.
Appendix C

Optimality System for HIV Model

Recalling our model equations (4.1) and our optimal cost (4.18) then we can write the Hamiltonian for our problem as

\[
H(T_1, T_2, T_1^*, T_2^*, V, V_{NI}, E) = Q_eV(t) + R_1\epsilon^2_1(t) + R_2\epsilon^2_2(t) - SE(t) \\
+ p_1(\lambda_1 - d_1T_1 - (1 - \epsilon_1)k_1VIT_1) \\
+ p_2(\lambda_2 - d_2T_2 - (1 - f\epsilon_1)k_2VIT_2) \\
+ p_3((1 - \epsilon_1)k_1VIT_1 - \delta T_1^* - m_1ET_1^*) \\
+ p_4((1 - \epsilon_1)k_2VIT_2 - \delta T_2^* - m_2ET_2^*) \\
+ p_5((1 - \epsilon_2)N_T\delta(T_1^* + T_2^*) - [c + (1 - \epsilon_1)\rho_1k_1T_1 + (1 - f\epsilon_1)\rho_2k_2T_2]V(t) \\
+ p_6(\epsilon_2N_T\delta(T_1^* + T_2^*) - cV_{NI}) \\
+ p_7(\lambda E + b_E\frac{T_1^* + T_2^*}{T_1^* + T_2^* + K_b}E - d_E\frac{T_1^* + T_2^*}{T_1^* + T_2^* + K_d}E - \delta_E E).
\]  

(C.1)

Now we will differentiate \( H \) with respect to the state variables \( T_1, T_2, T_1^*, T_2^*, V, V_{NI}, \) and \( E \) to get the equations for the costate equations. Or, mathematically,

\[
-\dot{p}_i = \frac{\partial H}{\partial x_i},
\]  

(C.2)
where $x = [T_1\ T_2\ T_1^*\ T_2^*\ V\ E]^T$. In particular,

$$\dot{p}_1 = -(p_1(-d_1 - (1 - \epsilon_1)k_1V) + p_3(1 - \epsilon_1)k_1V - p_5(1 - \epsilon_1)\rho_1k_1V)$$

$$\dot{p}_2 = -(p_2(-d_2 - (1 - f\epsilon_1)k_2V) + p_4(1 - f\epsilon_1)k_2V - p_5(1 - f\epsilon_1)\rho_2k_2V)$$

$$\dot{p}_3 = -(p_3(-\delta - m_1E) + p_5(1 - \epsilon_2)N_T\delta$$

$$+ p_7\left(\frac{b_E EK_b}{(T_1^* + T_2^* + K_b)^2} - \frac{d_E EK_d}{(T_1^* + T_2^* + K_d)^2}\right) + p_6e_2N_T\delta)$$

$$\dot{p}_4 = -(p_4(-\delta - m_2E) + p_5(1 - \epsilon_2)N_T\delta$$

$$+ p_7\left(\frac{b_E EK_b}{(T_1^* + T_2^* + K_b)^2} - \frac{d_E EK_d}{(T_1^* + T_2^* + K_d)^2}\right) + p_6e_2N_T\delta)$$

$$\dot{p}_5 = -(Q_e - p_1(1 - \epsilon_1)k_1T_1 - p_2(1 - f\epsilon_1)k_2T_2 + p_3(1 - \epsilon_1)k_1T_1$$

$$+ p_4(1 - f\epsilon_1)k_2T_2 + p_5(-c - (1 - \epsilon_1)\rho_1k_1T_1 - (1 - f\epsilon_1)\rho_2k_2T_2))$$

$$\dot{p}_6 = cp_6$$

$$\dot{p}_7 = -(-S - p_3m_1T_1^* - p_4m_2T_2^*$$

$$+ p_7\left(\frac{b_E(T_1^* + T_2^*)}{T_1^* + T_2^* + K_b} - \frac{d_E(T_1^* + T_2^*)}{T_1^* + T_2^* + K_d} - \delta_E\right)).$$

Finally, the derivatives of the Hamiltonian with respect to the controls $\epsilon_1$ and $\epsilon_2$ are given by

$$\frac{\partial H}{\partial \epsilon_1} = 2R_1\epsilon_1 + (p_1 - p_3 + \rho_1p_5)k_1VT_1 + (p_2 - p_4 + \rho_2p_5)fk_2VT_2$$

$$\frac{\partial H}{\partial \epsilon_2} = 2R_2\epsilon_2 - p_5N_T\delta(T_1^* + T_2^*)p_6N_T\delta(T_1^* + T_2^*).$$

Note that by setting these derivatives equal to zero is generally described as the stationary condition. Finding the triple $x, u, \text{ and } p$ that satisfy the state equations (4.29), costate equations (C.3), and stationary equations (C.4) gives the solution to the optimal control problem. As our control is bounded, we replace this condition with Pontryagin’s minimum principle and use this derivative in a gradient based algorithm to approximate the solution that satisfies this condition.