

# COMPUTER BASED HIV/AIDS PARAMETER ESTIMATION FOR PRACTITIONERS AND PATIENTS

R. Filter <sup>\*,1</sup> X. Xia <sup>\*</sup>

*\* Department of Electrical, Electronic and Computer Engineering, University of Pretoria, Pretoria 0002, South Africa*

**Abstract:** This paper describes a computer application of parameter estimation for the three dimensional HIV/AIDS model. The program uses a least square based procedure with standard optimization routines and aims to allow parameter extraction for individual patients. Together with the basic parameter estimation, it is shown how additional information from outside a measurement dataset can be included in the estimation routine to increase the reliability and accuracy of parameter estimates. The developed application uses the theoretical basis of cost function based, parameter estimation to allow the medical practitioner to extract the three dimensional model parameters for HIV/AIDS patients. The program gives greater insight into the progression of the disease, and helps the practitioner to decide on the correct dosage for each patient. Also, advice on the time to initiate therapy, the drug dosing and the interval between tests is given.

**Keywords:** Computer based treatment advisor, medical systems, HIV/AIDS modelling, physical parameters and parameter estimation.

## 1. OVERVIEW AND BACKGROUND

Human Immunodeficiency Virus/Acquired Immunodeficiency Syndrome (HIV/AIDS) is an area of active research in numerable medical institutions. Mathematical modelling in combination with highly active antiretroviral therapy (HAART) resulted in a paradigm shift about the disease and its treatment. However, there are still many uncertainties in the area of HAART. The best time to initiate therapy, dosage levels and the optimal combination of drugs are areas currently demanding further research. Together with these, the side effects of different drugs add to the complexity of finding treatment strategies. There is thus a need to determine the influence of

HAART on the virus and on the immune system on an individual patient basis. This is necessary to insure that each patient can be treated according to personal needs.

A helpful tool to decide on dosages in treatment of HIV/AIDS is a basic model that describes the disease and the influence of drugs on the virus. A three-dimensional mathematical model has helped to reshape the perception of the disease, by allowing the estimation of key parameters, such as the half-lives of infected cells and free virus. The published estimates are for a subset of the parameters only, and give an indication of parameter values for a cohort of patients. In order to use this model as a tool for treatment decisions, it is necessary to determine all six parameters of the model for individual patients. Even though there are general observations that can be made from the model and

---

<sup>1</sup> Corresponding author. Phone: +27 (82) 442 3415 Fax: +27 (12) 362 5000 Email: s9801305@postino.up.ac.za

its structure, it is only when the model is tailored to each patient's individual parameters that clear benefits in the treatment strategy arise.

It is important that model parameters can be determined from measurements that are acquired on equipment that is accessible to local health services. This paper presents a computer-based program that can be used to extract all six of the model parameters from patient data, even under less favorable conditions.

Some measurements from clinics might contain enough information to extract useful parameters, even when it is not possible to extract all six parameters. In such situations the program can be used to accommodate generalizations of some parameters that do not vary considerably between patients.

The application is developed to allow the medical practitioner to estimate the parameters for HIV/AIDS patients. This allows greater insight into the progression of the disease, and helps the practitioner to decide on the correct dosage for the patient. The therapy scheduler gives advice on the time to initiate therapy, the drug dosing and the interval between tests.

The program can be used at different levels of complexity. The basic program demands no background knowledge from the user. For practitioners with experience in mathematical modelling, there is ample opportunity to fine-tune the procedures to conform to the special needs that may arise from different medical situations.

## 2. THEORETICAL BASIS

This section describes the approach of parameter estimation presented in (Filter and Xia, 2003). A three-dimensional model of HIV/AIDS, used in this program, consists of three variables: the population sizes of uninfected cells ( $T$ ), infected cells ( $T^*$ ), and free virus particles ( $v$ ). Free virus particles infect uninfected cells at a rate proportional to the product of their abundances,  $\beta vT$ . The rate constant,  $\beta$ , describes the efficacy of this process. Infected cells produce free virus particles at a rate proportional to their abundance,  $kT^*$ . Infected cells die at a rate  $\delta T^*$ , and free virus particles are removed from the system at a rate  $cv$ . By assuming a constant production rate,  $s$  and death rate  $dT$  for the uninfected cells, the three-dimensional model of virus dynamics is obtained (Nowak and May, 2000; Perelson and Nelson, 1998):

$$\begin{cases} \dot{T} = s - dT - \beta T v, \\ \dot{T}^* = \beta T v - \delta T^*, \\ \dot{v} = k T^* - c v. \end{cases} \quad (1)$$

Furthermore, for the purpose of estimating model parameters, it will be assumed that plasma viral load and CD4<sup>+</sup> T cell count are measured. That is, the measurement outputs are  $y_1 = T$ , and  $y_2 = v$ . This is in accordance with the current prevailing medical practice (Panel on Clinical Practices for Treatment of HIV Infection, 2001).

The basis of parameter estimation used in the program is the squared distance between measurement points and a trajectory of points generated with  $\hat{\chi}$ . As with the method considered by Xia (2002), this method is in essence least square (LSQ) based, but with two important differences. Firstly, derivative estimation is only present when a nominal curve is generated by a numerical ordinary differential equation (ODE) solver and, thus, this estimation is not influenced by measurement noise.

Secondly, the cost function is not limited to the LSQ distance, thus, it can be expanded to accommodate a diverse base of knowledge in order to increase the accuracy of parameter estimation.

A pre-existing implementation of the Nelder-Mead Simplex search method is used here as the optimization routine, to find a set of parameters that minimizes the cost function<sup>2</sup>. At each iteration of the search, the cost function is called by the optimization routine. When a pre-set tolerance is met by the optimization routine, it exits with the final parameter estimation. The basic steps of the cost function are as follows:

- (1) The function receives a list of data points for the CD4<sup>+</sup> T cell and the virus count with their respective time points. Together with these,  $\hat{\chi}$  is also passed to the function.
- (2) When constraints are specified for  $\chi$ , they are enforced at this point. (By default only positive values are allowed for the parameters.)
- (3) The function uses  $\hat{\chi}$  and solves the dynamic model of eq. (1). This is done within the framework of a pre-existing numerical ODE solver.
- (4) The numerical solution is used to calculate the difference between each data point and its predicted value. The differences are squared and summed.
- (5) Any additional penalty values are calculated and added to the total, which is then returned to the optimization function.

It should be clear that this method is not prone to the derivative estimation error, as is the LSQ method considered in (Xia, 2002).

---

<sup>2</sup> It should be emphasized that the estimation procedure is not dependent on the use of the Nelder-Mead search method. Other search methods could be used. It is important that the chosen search method does not rely on a smooth cost function.

In the rest of this section, the cost function will be discussed in more detail.

### 2.1 Pure LSQ with correction factor

Apart from the problem of derivative estimation, there is a second major drawback in the pure LSQ method. The equations that are fitted to the data, contain product terms of CD4<sup>+</sup> T cell and virus counts. This essentially forces the data vectors to be of equal length for proper estimation. If this is not the case, interpolation has to be employed, which degrades the results. Since the penalty function method does not require any product terms, there is no constraint on the length of CD4<sup>+</sup> T cell and virus data vectors. In fact none of the points of the two data vectors have to coincide in time. Thus, for a nominal set of parameters,  $\hat{\chi}$ , and initial conditions,  $\hat{\boldsymbol{x}}_0$ , a curve is generated with a numerical ODE solver to find  $\hat{T}$  and  $\hat{v}$ . Together with  $N$  measurements of  $T$  and  $K$  measurements of  $v$ , at time  $t_1, \dots, t_N$ , and  $\tau_1, \dots, \tau_K$  respectively, we define the basic cost function as,

$$J_w = \sum_{n=1}^N \frac{(\hat{T}(t_n) - T_n)^2}{N \text{mean}(T_n)} + \sum_{k=1}^K \frac{(\hat{v}(\tau_k) - v_k)^2}{K \text{mean}(v_k)}. \quad (2)$$

The differences in data points may not jeopardize the balance of the penalty function, thus both data vectors have to be weighted by their mean value and their length. For equal length data vectors, division by their length is not necessary.

### 2.2 Logarithmic distance

From (CDC Working group, 2001) it is known that the tests used to determine viral load are log based. Even with the highest precision tests, a log variance of 0.6 can be expected in the measurements. For the Roche<sup>®</sup> Amplicor HIV-1 Monitor<sup>™</sup> test, the most commonly used test by the participating laboratories in (CDC Working group, 2001), a log difference of up to 2.2 was noted. Also for the experiments performed by Perelson and Nelson (1998), the virus data is fitted to the least square of logarithmic distance.

For this case the cost function can be modified as follows,

$$J_l = \sum_{n=1}^N \frac{(\hat{T}(t_n) - T_n)^2}{\text{mean}(T_n)N} + \sum_{k=1}^K \frac{(\log \hat{v}(\tau_k) - \log v_k)^2}{\text{mean}(\log v_k)K}. \quad (3)$$

Thus, the logarithmic distance between virus data points is used in the least square calculation. In this example the CD4<sup>+</sup> T cell data term is still computed as a linear value. Similar changes can be made for this term if necessary.

### 2.3 Additional refinements

It is often the case that a set of data on its own does not contain enough information for a complete determination of parameters, but when the prevailing circumstances are known, this knowledge would allow the extraction of key parameters. In these situations the custom penalty function is helpful, since outside knowledge of the dataset can be incorporated into the parameter estimation cycle.

The main points where refinements can be incorporated are at point 2 and 5 in the penalty function. Some of the common refinements that are used in the program are described below.

#### Enforcing limits

This is usually done at step 2 in the penalty function, by checking the parameters against a predefined range and correcting any parameters that do not fall within the specifications. A second option would be to add these limits directly to the penalty value at step 5. This would allow for weighted penalties if any of the limits are violated.

#### Prior knowledge of parameters

When a parameter is known from another source (*e.g.*, experiment, assumption or literature), this knowledge is incorporated at step 2. At the same time the optimization routine has to be instructed not to search for the parameters that are already known.

#### Prevailing conditions

When prevailing conditions (*e.g.*,  $c > \delta$ ) for a dataset are known from other sources, these conditions are usually added by means of an additional term in step 5. This term must be scaled to ensure its proper influence.

As an example, consider the experiment described in (Perelson and Nelson, 1998, pp. 16–19). In this experiment key assumptions were made in order to extract two of the six parameters. Firstly each patient was assumed to be at steady state (“set-point” has been reached) before initiation of therapy. This is a prevailing condition for that experiment, since the author had access to viral load data before the experiment, which indicated that the viral loads were in steady state. Secondly Nowak and May (2000, p.32) state a prior knowledge that infected cells live longer than free virus. This information is reflected in  $J_r$  by adding two terms to the basic LSQ cost.

$$J_r = J + k_1 \max\left(\frac{d\hat{v}_s}{dt}, 0\right) + k_2 \max(\hat{\delta} - \hat{c}, 0), \quad (4)$$

where  $J$  is either  $J_w$  or  $J_l$ , and  $\hat{v}_s$  is the vector of computed viral loads, truncated after a few days. In this case scaling constants  $k_1$  and  $k_2$  are chosen such that any violation of the prevailing

conditions would result in a clear increase of the penalty function.

The first refinement term corresponds to the knowledge that the patient is in steady state before initiation of therapy. Thus, no positive derivative should be allowed initially, for the viral load. An intuitive way to see this is to note that therapy results in a decline of virions, thus, an increasing virion count could only be the result of fluctuations before therapy, which is not possible since the patient was in steady state at the start of therapy. The second refinement term corresponds to the statement that the average infected CD4<sup>+</sup> T cell lives longer than free virions.

Some tests were performed to ensure that this estimation procedure produces viable parameter values, with encouraging results. Interested readers are referred to (Filter and Xia, 2003) for more information.

### 3. FROM THEORY TO PRACTICE

By using the theoretical background as a springboard, a program is created that allows users from diverse backgrounds to increase the effectivity of HIV/AIDS treatment strategies. With the insights obtained from parameter estimation, users can monitor individual patients without the need to understand underlying mathematical details of the model.

Some of the key points for the user are listed below:

- The user interface allows both patients and general practitioners to enter their sample data and receive treatment guidelines. Data can be entered in the familiar environment of Microsoft Excel.
- Information falling outside the basic dataset can be entered and is incorporated into the estimation procedure to increase the accuracy of estimation.
- The program determines a set of initial conditions from the data and external information for the identification routine without the intervention of an HIV/AIDS modelling expert.
- Inconsistencies in results and data are detected and the user is informed about the discrepancies to allow expert help to be obtained.
- If treatment is initiated after an original set of data has been examined, the program can analyze the influence of treatment and give an indication of its effectiveness.
- In situations where a user needs detailed information, or where personal limits must be added that are not currently interpreted by

the program, a higher level user can modify the cost function directly and access the raw output from the identification routine.

Some of the underlying functions are listed below:

- The program follows a control-system approach to model identification using a modifiable cost function as described in section 2.

This is the key function of the program and in situations where enough data points are available this would be adequate for parameter estimation with a static cost function.

- The cost function is LSQ based and combines the basic LSQ cost with the bounds and prevailing conditions from the user supplied information.

Here the user information from outside the data set comes into play. As an example: When a patient supplies the program with an estimated time since infection that indicates that the steady state in viral load has been reached, the patient is advised to take medication in conjunction with the load tests. From the point where medication is taken, the cost function is modified as in section 2.3. Here, the advice to take medication is based on the necessary conditions for parameter estimation. For more information about the necessary conditions for parameter estimation see (Xia and Moog, 2003).

- The fine-tuning of the cost function is hidden from the user and user input about prevailing circumstances is translated into mathematical terms.

Direct user intervention in the cost function is not necessary for the previous point. Pre-programmed knowledge is incorporated automatically, with feedback to the user. If any assumption needs to be overridden, the user can select to do so without directly editing the cost function.

- A parameter search is done by standard means and the resulting parameters are interpreted for the user.

Once the basis of data and external information is in place, the parameter search commences along standard avenues. It is in the interpretation of results that there is still a great deal of knowledge necessary. Since the parameters are not yet in standard use among practitioners, there is no experience base from which to make recommendations. Currently, the recommendations are made according to general research results as given, for instance, in (Jeffrey *et al.*, 2003). The program allows theoretical scenarios to be viewed with different treatment strategies. As more individual experience is gained by practitioners and patients, this information

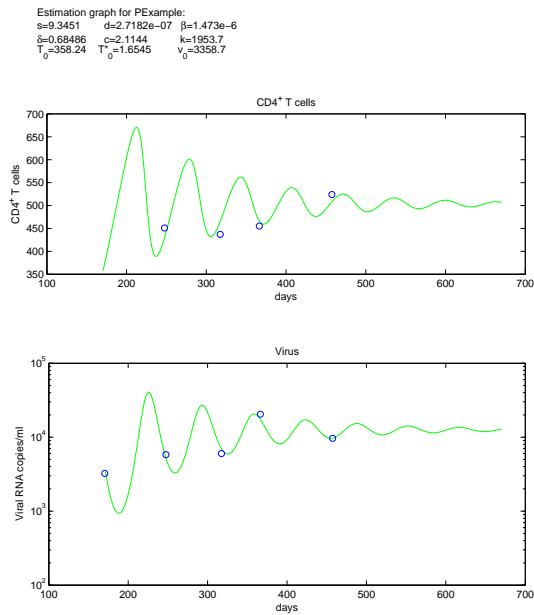


Fig. 1. Visual printout of estimated parameters for an example patient

can be used to generate more pointed recommendations.

- In situations where there is not enough data from a single patient, the program can use information from its database to give possible scenarios that allow the user to make initial treatment decisions.

Here the user can benefit from general parameter estimates of other studies like (Perelson and Nelson, 1998). Since there is no detailed data available for the individual, some parameters can be fixed at expected values to generate a general scenario that contains those initial data points and considers the external information. If, for instance, the patient has been infected for more than fifteen months, it is likely that viral load is in steady state (Filter and Xia, 2003), and from a single sample some of the ratios of parameters can be estimated. Even if the finer details are not available, the user still gets an indication of possible scenarios.

As an example, the estimation for a patient is shown in figure 1, with the corresponding report for this patient shown in figure 2. The report is text based and contains the main information and advice. The user can also see a graphical representation of the parameters in a treatment context and compare different scenarios of treatment as shown in figure 3. The scenarios represented in this figure are based purely on theoretical values of drug efficacy. Since patients can be monitored after the initiation of therapy, these theoretical values can be augmented by the information gained during treatment. Thus, the trajectory of viral load after initiation of therapy may be used to

PExample:  
 Basis Report  
 -----  
 Enough data points were available to determine parameters:  
 $s=11.409$   
 $d=0.008$   
 $\beta=1.2 \times 10^{-6}$   
 $\delta=1.35$   
 $c=1.31$   
 $k=2889.2$   
 Set-point= 12223  
 Estimated time to reach set-point = 14 months

Constraints on parameters were set to default.  
 $c < \delta$   
 $s < 50$   
 $d > 0.008$   
 No additional modifications made cost function

Fit seems adequate: no outliers were detected in data-set.

The spacing between data points is very large, it is important to perform a visual check on the data to ensure that a good fit has been found. It is recommended that samples are taken closer together. For more details see the guidelines for sample spacing.

CD4 cell count seems stable at an acceptable level. The viral set-point falls in the lower range: according to the panel on clinical practices on HIV/AIDS it might be desirable to postpone a rigorous treatment regime.

For visual scenarios of different treatment strengths, please consult the treatment advisor application.

Fig. 2. Basis report printout of estimated parameters for an example patient

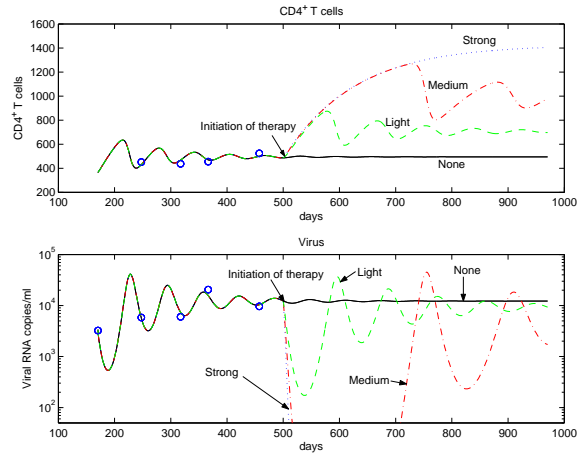


Fig. 3. Treatment advisor visual printout for different efficiencies of protease inhibitors

determine the efficacy of a regime even before viral load is suppressed below the benchmark values described by the Panel on Clinical Practices for Treatment of HIV Infection (2001).

The described program is in the early stages of development where only the key functions are implemented. In order to keep the program relevant, the advice generator must be kept up to date in accordance with current guidelines and experience gained from program use. One of the future development goals is to incorporate parameter estimation for higher-dimensional models for patients with enough data points.

It should be stressed that a computer program cannot replace the expert care of a medical practitioner. Program users should always be aware that this program is a tool to help in decision-

making and planning of treatment, but that the final decisions lie with the medical practitioner in dialogue with the patient.

#### 4. CONCLUSION

The developed application uses the theoretical basis of cost function based parameter estimation to allow the medical practitioner to estimate the three dimensional model parameters for HIV/AIDS patients. The program gives greater insight into the progression of the disease, and helps the practitioner decide on the correct dosage for the patient.

Even though these basic functions are implemented, the advice of the program is currently based on a small experience base. As the individually tailored parameter estimates become a part of each patients treatment strategy, the experience of practitioners in the field and insights from research can be incorporated to increase the quality of advice given by the program.

In order to cater for personal experience in treatment strategies, the program can be used at different levels of complexity. The basic program demands little background knowledge from the user, but for practitioners with experience in mathematical modelling, there is ample opportunity to fine-tune the procedures to conform to the special needs that may arise from their medical situations.

#### REFERENCES

- CDC Working group (2001). "Guidelines for Laboratory Test Result Reporting of Human Immunodeficiency Virus Type 1 Ribonucleic Acid Determination". <http://www.cdc.gov/>.
- Filter, R. A. and X. Xia (2003). A penalty function approach to HIV/AIDS model parameter estimation. 13th IFAC Symposium on System Identification. Rotterdam, Netherlands.
- Jeffrey, A. M., X. Xia and I.K. Craig (2003). When to initiate HIV therapy: A control theoretic approach. *IEEE Transactions on Biomedical Engineering*. (to appear).
- Nowak, Martin A. and Robert M. May (2000). *Virus Dynamics: Mathematical Principles of Immunology and Virology*. Oxford University Press.
- Panel on Clinical Practices for Treatment of HIV Infection (2001). "Guidelines for the Use of Antiretroviral Agents in HIV-Infected Adults and Adolescents". <http://www.hivatis.org>.
- Perelson, Alan S. and Patrick W. Nelson (1998). Mathematical analysis of HIV-1. *SIAM Review* **41**(1), 3-44.

Xia, X. (2002). Estimation of HIV/AIDS parameters. IFAC. 15<sup>TH</sup> IFAC World Congress, Barcelona, Spain. to appear in *Automatica*, 2003.

Xia, X. and C. H. Moog (2003). Identifiability of non-linear systems with applications to hiv/aids models. *IEEE Transactions on Automatic Control*. (to appear).