

IDENTIFICATION OF HIV-1 DYNAMICS

Estimating the Noise Model, Constant and Time-varying Parameters of Long-term Clinical Data

Andras Hartmann^{1,2}, Susana Vinga^{1,3} and Joao M. Lemos^{1,2}

¹INESC-ID, R. Alves Redol 9, 1000-029 Lisboa, Portugal

²IST-UTL, Av. Rovisco Pais, 1049-001 Lisboa, Portugal

³FCM-UNL, C. Martires Patria 130, 1169-056 Lisboa, Portugal

Keywords: HIV-1 viral dynamics, Parameter identification, Non-linear, Differential equation, Time-varying parameter.

Abstract: The importance of a system theory based approach in understanding immunological diseases, in particular the HIV-1 infection, is being increasingly recognized. This is because the dynamics of virus infection may be effectively represented by relatively compact state space models in the form of nonlinear ordinary differential equations. This work focuses on the identification of constant and time-varying parameters in long-term dynamic HIV-1 data. We introduce a novel strategy for parameter identification. Constant parameters were estimated using Particle Swarm Optimization (PSO), and time-varying parameters were captured with Extended Kalman Filter (EKF). As EKF relies on the noise strongly, the measurement noise was also inferred. The results are convincing on clinical data: similar noise parameters were detected for two different subjects, a good overall fit was reached to the data, and EKF was found efficient in estimating the time-varying parameters, overcoming drawbacks and limitations of existing methods.

1 INTRODUCTION

The exhaustive study of HIV viral dynamics since the early 90's has lead to a deeper insight into the pathology of the infection (Perelson et al., 1996). Ordinary differential equations (ODEs) were introduced as a powerful tool to describe the underlying processes (Perelson and Nelson, 1999). Of course, well established models and identification algorithms are essential, which are intensively developed, see the comprehensive review (Wu, 2005) and references therein. The importance of long-term dynamics in HIV infection modeling was just recently recognized. In the early models parameters were considered as being constant, which is a good approximation for the short-term behavior, however in long-term dynamics some (if not all) parameters may change over time due to variation in treatment effects. Time-varying parameters for drug adherence were studied in (Huang et al., 2003; Huang, 2008). Recently, (Liang et al., 2010) claimed to be the first to estimate both constant and time-varying parameters from the time-series only, not using historical parameters from prior studies in any sense, nor similarity with other time-series.

The contribution of the present paper consists in

suggesting a novel strategy for more flexible identification of time-varying parameter by avoiding splines and without introducing new parameters to the model. As the method relies on the statistical noise model, we also characterized the noise contaminating the data, which was found to be multiplicative zero mean Gaussian.

2 METHODS

2.1 Dataset

The longitudinal clinical dataset used here was previously introduced by (Liang et al., 2010) for parameter estimation. It contains long-term viral load (copies/ml) together with CD4+ T cell counts (copies/mm³) data of two patients. The infected and uninfected cells are measured together, thus only the total CD4+ T count ($T + T^*$) is accessible. The smallest detectable viral load with the available technology is about 50 copies/ml. In the original dataset the values below were substituted with the threshold. In this study, the undetectable values were excluded.

2.2 Estimation of the Noise Model

To estimate measurement noise, the time-series were smoothed with least squares estimates regularized up to 4th order local derivatives as described in (Liang et al., 2010). To ensure that the noise could come from independent identical normal distribution, normality of the noise was tested with two different methods: Chi-square and Lilliefors test. Gaussian noise model has the advantage that it is easy to deal with, and has only the first two moments (mean and variance) as parameters, which could be determined from the dataset. However central limit theorem suggests that the noise should be Gaussian, to the best knowledge of the authors this has never been tested empirically before.

2.3 The Dynamic Model of the HIV-1 Infection

Although many, more complex parametric ODE models of the HIV infection can be found in the literature, incorporating different aspects of the process (Wu, 2005), the most widely used one is the three dimensional *basic model* described by Eqs. (1-3) see e.g. (Perelson and Nelson, 1999). This is a consequence of the reduced datasets available and the fact that in many cases data is collected with a very low sampling rate, suggesting the usage of models as simple as possible.

$$\dot{T} = s - dT - \beta T v \quad (1)$$

$$\dot{T}^* = \beta T v - \xi T^* \quad (2)$$

$$\dot{v} = k T^* - c v \quad (3)$$

The model includes three state variables: the concentration of healthy CD4+ T cells ($T = T(t)$), infected CD4+ T cells ($T^* = T^*(t)$), and free virus particles ($v = v(t)$). Healthy CD4+ T cells are produced at a constant rate s , and have an average life span of $1/d$ days. These cells can be infected by free virus particles, and become infected cells. The infection is modeled using a simple mass-action type term, with a rate constant β . Infected cells may have a different life span ($1/\xi$) than healthy cells, which means that in general $\xi \neq d$. Finally, free virus particles are produced in infected cells, and released at a rate k , having an average life span of $1/c$.

The basic model was subject to several identifiability studies, where all constant and time-varying parameters were found structurally (mathematically) identifiable in case the initial conditions are known (Wu et al., 2008; Liang et al., 2010).

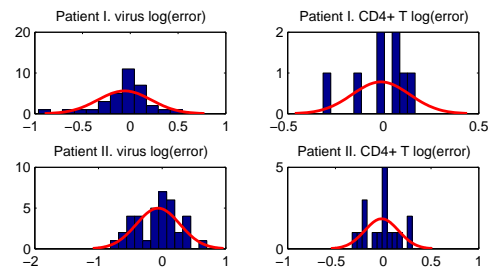


Figure 1: Fitting Gaussian to the estimated noise.

2.4 Parameter Estimation

Parameter estimation is the process where the best fitting parameters (given an objective function) are determined. First, all parameters of the ODE system in Eqs. (1-3) were treated as being constant, and identified with PSO (Kennedy et al., 1995). This resulted in an initial parameter set, which was further refined by introducing the time-dependency to the infection rate (β). For the estimation of the time-varying parameter, a continuous-discrete version of EKF (CD-EKF) (Sarkka, 2006) was applied. The continuous ODE model together with the initial parameter set and the noise model was plugged in into the filter. The time-varying parameters were estimated from the discrete measurements at hand. The time update between the measurements was simulated by solving the differential equations, while a measurement update of the standard EKF was executed any time a measurement was available. This approach was introduced in (Sarkka, 2006; Kristensen, 2004), and handles both irregular sampling and missing data problems. All the noise was considered to be measurement noise, while process noise was set to zero.

3 RESULTS

The noise on the virus concentrations shows exponential-like decay, and thus the log-errors were plotted in Figure 1. Note that for patients with treatment, the virus concentration also has an exponential decay, and in the earliest models even exponentials were fitted to the data (Perelson et al., 1996; Wu, 2005). The visual fitting of the noise on the viral load indicates that it may be approximated with Gaussian. On the other hand, the number of data instances in CD4+ T data is insufficient to reach a conclusion upon visualization only, and it is even hardly enough to test against a distribution numerically. The results of the numerical normality tests in Table 1 confirm in most cases that the hypothesis of normality is not rejected. Chi-square test indicates that the errors of the viral

Table 1: Results and P-values of normality tests on the estimated noise. These tests include the null-hypothesis that the data has normal distribution with the outcome being 1 if the null hypothesis can be rejected at the 5% significance level (P-value<0.05), and 0 otherwise. The P-value=NaN indicates that it could be determined due to insufficient amount of data.

	Patient I				Patient II			
	V		T + T*		V		T + T*	
	Result	P-value	Result	P-value	Result	P-value	Result	P-value
Chi-square	0	0.0661	0	NaN	0	0.7091	0	NaN
Lilliefors test	1	0.0108	0	0.1821	0	0.4925	0	0.4928

Table 2: Sample mean (μ) and standard deviation (σ) of the noise.

	Patient I		Patient II	
	T	V	T	V
μ	-0.0599	-0.0118	-0.0623	-0.0163
σ	0.2790	0.1503	0.3422	0.1768

load are Gaussian, while it failed to determine the fit on the CD4+ T error because of insufficient amount of data. The Lilliefors test considers all the errors to be Gaussian except the one corresponding to the viral load of Patient I. The empirical parameters of the noise were found to be similar for the two patients and are listed in Table 2. According to these results, in the rest of the paper we will consider multiplicative zero mean Gaussian noise.

There is some variation among the discovered constant parameters, but they are of the same magnitude, see Table 3. For comparison purposes parameters estimated for the same time-series by (Liang et al., 2010) are also represented. Figure 2 shows reconstruction using constant parameters discovered by PSO. Figure 3 shows the time-varying parameter estimated by EKF using the noise parameters in Table 2.

4 DISCUSSION

By smoothing the data in the mean-squared sense we could reach the conclusion that the noise model should be rather multiplicative than additive. Even if in some cases the applied normality tests suffer from insufficient amount of data, mostly they implied that a zero mean Gaussian distribution is a good approximation of the noise. The fact that the noise parameters (sample mean and standard deviation) for two different time-series were found to be similar, suggests that the estimation procedure was appropriate and the noise model is valid. Note that, for a solid proof it should be confirmed on a larger, statistically more significant dataset with possibly more measurement instances.

Similarly to other previous studies like (Huang,

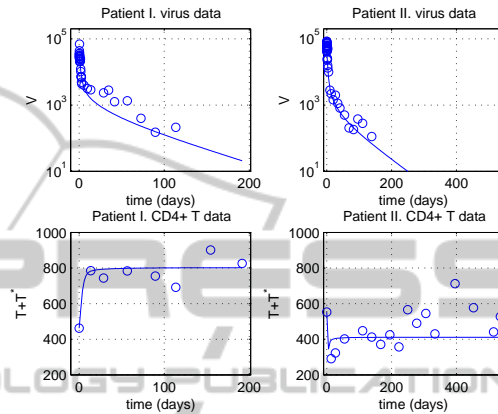


Figure 2: Reconstruction using constant parameters.

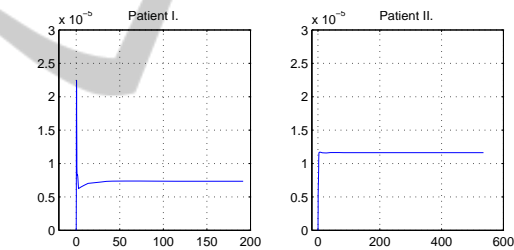


Figure 3: Time-varying parameter estimated by CD-EKF.

2008), inter-subject variability was observed in terms of parameters. It is worth noting that the discovered constant parameters with PSO are of the same magnitude, while parameters revealed by the MSSB and SNLS algorithms respectively show much larger deviations, (see Table 3). The differences between the parameter estimates of MSSB and PSO are also noticeable. This can be due to the different algorithms or the differences between the initial values, that can lead to completely different results (Wu, 2005). Another difference is that in this study we excluded the non-detectable values, while in (Liang et al., 2010) fitting was made to these values as well.

As the noise could be treated as zero mean Gaussian, the CD-EKF algorithm was a suitable choice to estimate the time-varying parameter: the Kalman-gain forces the parameter at every step towards its most likely value. Even with different constant pa-

Table 3: Estimated constant parameters with PSO compared with the results of the multistage smoothing-based (MSSB) approach and the spline-enhanced nonlinear least squares (SNLS) approach, see (Liang et al., 2010).

	Patient I			Patient II		
	PSO	MSSB	SNLS	PSO	MSSB	SNLS
s	368.94	254.49	397.09	238.40	18.38	45.45
d	0.46	0.34	0.49	0.58	0.04	0.10
ξ	2.16	1.09	1.09	2.48	0.43	0.43
k	1317.35	1284.27	288.57	1283.17	1190.81	479.18
c	3.60	2.46	2.46	2.50	3.78	3.78

parameters, the time-varying parameter shows similar behavior to those discovered with SNLS: After an initial perturbation they remain mostly constant. Patient I data shows much more significant overshoot in β than patient II, this can be due the different treatment effects, but it is also possible that this is an artifact of the method (e.g. numerical instability). Unlike SNLS algorithm we avoided the usage of splines for parameter estimation, that would address the design questions: what kind of spline and of which order to choose, moreover it would introduce new parameters.

5 CONCLUSIONS AND FUTURE WORK

Here we introduced a novel strategy for the estimation of constant and time-varying parameters of the long-term HIV-1 dynamic time-series. As this methodology depends on the noise, first the noise model was estimated from the data. The noise was found to be multiplicative zero mean Gaussian. Constant parameters were then estimated with PSO, while the time-varying parameter was further refined using an EKF algorithm adopted to continuous models. In comparison with existing methods here we found much smaller deviations between the estimated parameters, on the two patients. No approximation with smoothing or splines were used, instead, with EKF we applied the noise model directly to estimate the time-varying parameter in a Bayesian manner. Thus we believe this method offers a simpler and more flexible framework for estimating the time-varying parameter.

Our future work aims at better characterization of the noise and the time-varying parameter involving more patients' data and simulated experiments.

ACKNOWLEDGEMENTS

This work was supported by FCT (Portugal) under project PTDC/EEA-CRO/100128, HIVControl -

Control based on dynamic modeling of HIV-1 infection for therapy design, and INESC-ID multi-annual funding through the PIDDAC program funds. Andr as Hartmann received PhD fellowship from FCT - Fundao para a Ci ncia e a Tecnologia under the reference SFRH/BD/69336/2010.

REFERENCES

- Huang, Y. (2008). Long-term HIV dynamic models incorporating drug adherence and resistance to treatment for prediction of virological responses. *Computational Statistics & Data Analysis*, 52(7):3765–3778.
- Huang, Y., Rosenkranz, S. L., and Wu, H. (2003). Modeling HIV dynamics and antiviral response with consideration of time-varying drug exposures, adherence and phenotypic sensitivity. *Mathematical Biosciences*, 184(2):165–186.
- Kennedy, J., Eberhart, R., and Others (1995). Particle swarm optimization. In *Proceedings of IEEE international conference on neural networks*, volume 4, pages 1942–1948. Perth, Australia.
- Kristensen, N. (2004). Parameter estimation in stochastic grey-box models. *Automatica*, 40(2):225–237.
- Liang, H., Miao, H., and Wu, H. (2010). Estimation of constant and time-varying dynamic parameters of HIV infection in a nonlinear differential equation model. *The annals of applied statistics*, 4(1):460–483.
- Perelson, A., Neumann, A., Markowitz, M., Leonard, J., and Ho, D. (1996). HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. *Science*, 271(5255):1582.
- Perelson, A. S. and Nelson, P. W. (1999). Mathematical Analysis of HIV-1 Dynamics in Vivo. *SIAM Review*, 41(1):3.
- Sarkka, S. (2006). *Recursive Bayesian inference on stochastic differential equations*. Dissertation for the degree of doctor of science in technology, Helsinki University of Technology (Espoo, Finland).
- Wu, H. (2005). Statistical methods for HIV dynamic studies in AIDS clinical trials. *Statistical methods in medical research*, 14(2):171–92.
- Wu, H., Zhu, H., Miao, H., and Perelson, A. S. (2008). Parameter identifiability and estimation of HIV/AIDS dynamic models. *Bulletin of mathematical biology*, 70(3):785–799.