Parameter Selection Methods in Inverse Problem Formulation

H. T. Banks[1], Ariel Cintrón-Arias[1,2], and Franz Kappel[1,3]
[1]Center for Research in Scientific Computation
Center for Quantitative Sciences in Biomedicine
Department of Mathematics
North Carolina State University
Raleigh, NC 27695-8212
[2]Department of Mathematics and Statistics
East Tennessee State University
Johnson City, TN 37614-0663
and
[3]Institute for Mathematics and Scientific Computation
University of Graz
Graz, Austria A8010

November 3, 2010

Abstract

We discuss methods for *a priori* selection of parameters to be estimated in inverse problem formulations (such as Maximum Likelihood, Ordinary and Generalized Least Squares) for dynamical systems with numerous state variables and an even larger number of parameters. We illustrate the ideas with an in-host model for HIV dynamics which has been successfully validated with clinical data and used for prediction and a model for the reaction of the cardiovascular system to an ergometric workload.

Key Words: Parameter selection, inverse problems, sensitivity, Fisher Information Matrix, HIV models, cardiovascular modeling.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
03 NOV 2010		**00-00-2010 to 00-00-2010**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Parameter Selection Methods in Inverse Problem Formulation	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **North Carolina State University,Center for Research in Scientific Computation,Department of Mathematics,Raleigh,NC,27695-8212**	8. PERFORMING ORGANIZATION REPORT NUMBER **CRSC-TR10-03**
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
We discuss methods for a priori selection of parameters to be estimated in inverse problem formulations (such as Maximum Likelihood, Ordinary and Generalized Least Squares) for dynamical systems with numerous state variables and an even larger number of param- eters. We illustrate the ideas with an in-host model for HIV dynamics which has been successfully validated with clinical data and used for prediction and a model for the reaction of the cardiovascular system to an ergometric workload.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**29**	

1 Introduction

There are many topics of great importance and interest in the areas of modeling and inverse problems which are properly viewed as essential in the use of mathematics and statistics in scientific inquiries. A brief, noninclusive list of topics include the use of traditional sensitivity functions (TSF) and generalized sensitivity functions (GSF) in experimental design (what type and how much data is needed, where/when to take observations) [9, 10, 11, 16, 57], choice of mathematical models and their parameterizations (verification, validation, model selection and model comparison techniques) [7, 12, 13, 17, 21, 22, 24, 25, 41], choice of statistical models (observation process and sampling errors, residual plots for statistical model verification, use of asymptotic theory and bootstrapping for computation of standard errors, confidence intervals) [7, 14, 30, 31, 55, 56], choice of cost functionals (MLE, OLS, WLS, GLS, etc.) [7, 30], as well as parameter identifiability and selectivity. There is extensive literature on each of these topics and many have been treated in surveys in one form or another ([30] is an excellent monograph with many references on the statistically related topics) or in earlier lecture notes [7].

We discuss here an enduring major problem: selection of those model parameters which can be readily and reliably (with quantifiable uncertainty bounds) estimated in an inverse problem formulation. This is especially important in many areas of biological modeling where often one has large dynamical systems (many state variables), an even larger number of unknown parameters to be estimated and a paucity of longitudinal time observations or data points. As biological and physiological models (at the cellular, biochemical pathway or whole organism level) become more sophisticated (motivated by increasingly detailed understanding – or lack thereof – of mechanisms), it is becoming quite common to have large systems (10-20 or more differential equations), with a plethora of parameters (25 – 100) but only a limited number (50 – 100 or fewer) of data points per individual organism. For example, we find models for the cardiovascular system [16, Chapter 1] (where the model has 10 state variables and 22 parameters) and [51, Chapter 6] (where the model has 22 states and 55 parameters), immunology [49] (8 states, 24 parameters), metabolic pathways [32] (8 states, 35 parameters) and HIV progression [8, 43]. Fortunately, there is a growing recent effort among scientists to develop quantitative methods based on sensitivity, information matrices and other statistical constructs (see for example [9, 10, 11, 23, 28, 37, 38, 60]) to aid in identification or parameter estimation formulations. We discuss here one approach using sensitivity matrices and asymptotic standard errors as a basis for our developments. To illustrate our discussions, we will use two models from the biological sciences: a) a recently developed in-host model for HIV dynamics which has been successfully validated with clinical data and used for prediction [4, 8]; b) a global non-pulsatile model for the cardiovascular system which has been validated with data from bicycle ergometer tests [45, 16].

The topic of system and parameter identifiability is actually an old one. In the context of parameter determination from system observations or output it is at least forty years old and has received much attention in the peak years of linear system and control theory in the investigation of observability, controllability and detectability [6, 18, 19, 33, 39, 44, 47, 53, 54]. These early investigations and results were focused primarily on engineering applications, although much interest in other areas (e.g., oceanography, biology) has prompted more recent inquiries for both linear and nonlinear dynamical systems [5, 15, 29, 35, 42, 48, 59, 60, 61, 62].

2 Statistical Models for the Observation Process

One has errors in any data collection process and the presence of these errors is reflected in any parameter estimation result one might obtain. To understand and treat this, one usually specifies a *statistical model* for the observation process in addition to the *mathematical model* representing the dynamics. To illustrate ideas here we use ordinary least squares (OLS) consistent with an error model for absolute error in the observations. For a discussion of other frameworks (maximum likelihood in the case of known error distributions, generalized least squares appropriate for relative error models) see [7].

In order to be more specific we assume that the dynamics of the system is modeled by a system of ordinary differential equations:

$$\dot{x}(t) = g(t, x(t), \theta),\ t \geq t_0, \quad x(t_0) = x_0(\theta), \tag{1}$$

$$z(t) = h(t, x(t), \theta), \quad t \geq t_0,\ \theta \in \mathcal{A}, \tag{2}$$

where $G \subset \mathbb{R}^n$ and $\mathcal{A} \subset \mathbb{R}^p$ are open sets and $g : [t_0, \infty) \times G \times \mathcal{A} \to \mathbb{R}^n$, $x_0 : \mathcal{A} \to \mathbb{R}^n$ and $h : [t_0, \infty) \times G \times \mathcal{A} \to \mathbb{R}$ are sufficiently smooth functions. The set $\mathcal{A}$ is called the set of *admissible parameters* and $z(\cdot)$ is the *measurable output* of the system, which for simplicity we assume to be scalar. Let $x(t) = x(t; \theta)$ denote the solution of (1) for given $\theta \in \mathcal{A}$ and set

$$f(t, \theta) = h(t, x(t; \theta), \theta), \quad t \geq t_0,\ \theta \in \mathcal{A}.$$

Then

$$z(t) = f(t; \theta), \quad t \geq t_0,\ \theta \in \mathcal{A}, \tag{3}$$

is the output model corresponding to the model (1), (2). It is clear that an output model of the form (3) can also originate from dynamical models, where instead of the ODE-system (1) we may have a system of delay equations or some partial differential equation. In order to describe the observation process we assume there exists a vector $\theta_0 \in \mathcal{A}$, referred to as the *true or nominal parameter vector*, for which the output $z(t) = f(t, \theta_0)$ describes the output of the real system exactly. At given *sampling times*

$$t_0 \leq t_1 < \cdots < t_N,$$

we have measurements y_j for the output of the real system, $j = 1, \ldots, N$. It is also reasonably assumed that each of the N longitudinal measurements y_j is affected by random deviations ϵ_j from the true underlying output. That is, we assume that the measurements are given by

$$y_j = f(t_j; \theta_0) + \epsilon_j, \quad j = 1, \ldots, N. \tag{4}$$

The measurement errors ϵ_j are assumed to be realizations of random variables $\mathcal{E}_j$ satisfying the following assumptions:

(i) the errors $\mathcal{E}_j$ have mean zero, $\mathsf{E}(\mathcal{E}_j) = 0$;

(ii) the errors $\mathcal{E}_j$ have finite common variance, $\mathrm{var}(\mathcal{E}_j) = \sigma_0^2 < \infty$;

(iii) the errors $\mathcal{E}_j$ are independent (i.e., $\mathrm{cov}(\mathcal{E}_j, \mathcal{E}_k) = 0$ whenever $j \neq k$) and identically distributed.

According to (4) the measurements y_j are realizations of random variables Y_j, the *observations* at the sampling times t_j. Then the statistical model for the scalar observation process is

$$Y_j = f(t_j;\theta_0) + \mathcal{E}_j, \quad j = 1,\dots,N. \tag{5}$$

Assumptions (i) – (iii) imply that the mean of the observations is equal to the model output for the nominal parameter vector, $\mathsf{E}\big(Y_j\big) = f(t_j;\theta_0)$, and the variance in the observations is constant in time, $\mathrm{var}(Y_j) = \sigma_0^2$, $j = 1,\dots,N$.

For given measurements $y = (y_1,\dots,y_N)^{\mathsf{T}}$ the estimate $\hat{\theta}_{\mathrm{OLS}}$ for θ_0 is obtained by minimizing

$$J(y,\theta) = \sum_{j=1}^{N}\big(y_j - f(t_j;\theta)\big)^2 = |y - F(\theta)|^2 = |F(\theta) - F(\theta_0) - \epsilon|^2, \tag{6}$$

where we have set

$$F(\theta) = \big(f(t_1;\theta),\dots,f(t_N;\theta)\big)^{\mathsf{T}}, \quad \epsilon = \big(\epsilon_1,\dots,\epsilon_N\big)^{\mathsf{T}},$$

and $|\cdot|$ is the Euclidean norm in $\mathbb{R}^N$. The estimate $\hat{\theta}_{\mathrm{OLS}}$ is a realization of a random variable, the *least squares estimator* Θ_{OLS}. In order to indicate the dependence on N we shall write $\hat{\theta}^N_{\mathrm{OLS}}$ and Θ^N_{OLS} when needed. From [55] we find that under a number of regularity and sampling conditions, as $N \to \infty$, Θ^N_{OLS} is approximately distributed according to a multivariate normal distribution, i.e.,

$$\Theta^N_{\mathrm{OLS}} \sim \mathcal{N}_p\left(\theta_0, \Sigma_0^N\right), \tag{7}$$

where $\Sigma_0^N = \sigma_0^2\big(N\Omega_0\big)^{-1} \in \mathbb{R}^{p\times p}$ and

$$\Omega_0 = \lim_{N\to\infty}\frac{1}{N}\chi^N(\theta_0)^{\mathsf{T}}\chi^N(\theta_0).$$

The $N \times p$ matrix $\chi^N(\theta)$ is known as the *sensitivity matrix* of the system, and is defined as

$$\chi^N(\theta_0) = \Big(\frac{\partial f(t_i;\theta_0)}{\partial\theta_j}\Big)_{1\le i\le N,\ 1\le j\le p} = \frac{\partial F}{\partial\theta}(\theta_0) = \nabla_\theta F(\theta_0). \tag{8}$$

Asymptotic theory [7, 30, 55] requires existence and non-singularity of Ω_0. The $p \times p$ matrix Σ_0^N is the covariance matrix of the estimator Θ^N.

If the output model (3) corresponds to the model (1), (2) then the derivatives of f with respect to the parameters are given by

$$\frac{\partial f}{\partial\theta_j}(t,\theta) = \frac{\partial h}{\partial x}(t,x(t;\theta),\theta)\frac{\partial x}{\partial\theta_j}(t;\theta) + \frac{\partial h}{\partial\theta_j}(t,x(t;\theta),\theta), \quad j = 1,\dots,p, \tag{9}$$

where $w(t;\theta) = \big((\partial x/\partial\theta_1)(t;\theta),\dots,(\partial x/\partial\theta_p)(t;\theta)\big) \in \mathbb{R}^{N\times p}$ is obtained by solving

$$\begin{aligned}
&\dot{x}(t;\theta) = g(t,x(t;\theta),\theta), \quad x(t_0;\theta) = x_0(\theta),\\
&\dot{w}(t,\theta) = \frac{\partial g}{\partial x}\big(t,x(t;\theta),\theta\big)w(t;\theta) + \frac{\partial g}{\partial\theta}(t,x(t;\theta),\theta), \quad w(t_0;\theta) = \frac{\partial x_0}{\partial\theta}(\theta),
\end{aligned} \tag{10}$$

from $t = t_0$ to $t = t_N$. One could alternatively obtain the sensitivity matrix using difference quotients (usually less accurately) or by using automated differentiation software (for additional details on sensitivity matrix calculations see [7, 9, 27, 28, 34, 36]).

The estimate $\hat{\theta}_{\rm OLS} = \hat{\theta}^N_{\rm OLS}$ is a realization of the estimator $\Theta_{\rm OLS}$, and is calculated using a realization $\{y_i\}_{i=1}^N$ of the observation process $\{Y_i\}_{i=1}^N$, while minimizing (6) over θ. Moreover, the estimate $\hat{\theta}_{\rm OLS}$ is used in the calculation of the sampling distribution for the parameters. The generally unknown error variance σ_0^2 is approximated by

$$\hat{\sigma}^2_{\rm OLS} = \frac{1}{N-p}\sum_{j=1}^{N}\big(y_j - f(t_j;\hat{\theta}^N_{\rm OLS})\big)^2, \tag{11}$$

while the covariance matrix Σ_0^N is approximated by

$$\hat{\Sigma}^N_{\rm OLS} = \hat{\sigma}^2_{\rm OLS}\Big(\chi^N(\hat{\theta}_{\rm OLS})^{\mathsf{T}}\chi^N(\hat{\theta}_{\rm OLS})\Big)^{-1}. \tag{12}$$

As discussed in [7, 30, 55] an approximate for the sampling distribution of the estimator is given by

$$\Theta_{\rm OLS} = \Theta^N_{\rm OLS} \sim \mathcal{N}_p(\theta_0, \Sigma_0^N) \approx \mathcal{N}_p(\hat{\theta}^N_{\rm OLS}, \hat{\Sigma}^N_{\rm OLS}). \tag{13}$$

Asymptotic standard errors can be used to quantify uncertainty in the estimation, and they are calculated by taking the square roots of the diagonal elements of the covariance matrix $\hat{\Sigma}^N_{\rm OLS}$, i.e.,

$$\mathsf{SE}_k(\hat{\theta}^N_{\rm OLS}) = \sqrt{(\hat{\Sigma}^N_{\rm OLS})_{k,k}}, \quad k = 1, \ldots, p. \tag{14}$$

Before describing the algorithm in detail and illustrating its use, we provide some motivation underlying the use of the sensitivity matrix $\chi(\theta_0)$ of (8) and the Fisher Information Matrix $\mathcal{F}(\theta_0) = \big(1/\sigma_0^2\big)\chi^{\mathsf{T}}(\theta_0)\chi(\theta_0)$. Both of these matrices play a fundamental role in the development of the approximate asymptotic distributional theory resulting in (13) and (14). Since a prominent measure of the ability to estimate a parameter is related to its associated standard errors in estimation, it is worthwhile to briefly outline the underlying approximation ideas in the asymptotic distributional theory.

Ordinary least squares problems involve choosing $\Theta = \Theta_{\rm OLS}$ to minimize the difference between observations Y and model output $F(\theta)$, i.e., minimize $|Y - F(\theta)|$. However the approximate asymptotic distributional theory (e.g., see [55, Chapter 12]) which is *exact* for model outputs linear in the parameters, employs a fundamental linearization in the parameters in a neighborhood of the hypothesized "true" parameters θ_0. Replacing the model output with a first order linearization about θ_0, we then may seek to minimize for θ in the approximate functional

$$|Y - F(\theta_0) - \nabla_\theta F(\theta_0)[\theta - \theta_0]|.$$

If we use the statistical model $Y = F(\theta_0) + \mathcal{E}$ and let $\delta\theta = \theta - \theta_0$, we thus wish to minimize

$$|\mathcal{E} - \chi(\theta_0)\delta\theta|,$$

where $\chi(\theta_0) = \nabla_\theta F(\theta_0)$ is the $N \times p$ sensitivity matrix defined in (8). This is a standard optimization problem [46, Section 6.11] whose solution can be given using the pseudo inverse $\chi(\theta_0)^\dagger$ defined in terms of minimal norm solutions of the optimization problem and satisfying $\chi(\theta_0)^\dagger = (\chi(\theta_0)^{\mathsf{T}}\chi(\theta_0))^\dagger\chi(\theta_0)^{\mathsf{T}} = \sigma_0^2\mathcal{F}(\theta_0)^\dagger\chi(\theta_0)^{\mathsf{T}}$. The solution is

$$\delta\Theta = \chi(\theta_0)^\dagger\mathcal{E}$$

or

$$\Theta_{\rm LIN} = \theta_0 + \chi(\theta_0)^\dagger\mathcal{E} = \theta_0 + \sigma_0^2\mathcal{F}(\theta_0)^\dagger\chi(\theta_0)^{\mathsf{T}}\mathcal{E}.$$

If $\mathcal{F}(\theta_0)$ is invertible, then the solution (to first order) of the OLS problem is

$$\Theta_{\mathrm{OLS}} \approx \Theta_{\mathrm{LIN}} = \theta_0 + \sigma_0^2 \mathcal{F}(\theta_0)^{-1} \chi(\theta_0)^{\mathsf{T}} \mathcal{E}. \tag{15}$$

This approximation, for which the asymptotic distributional theory is exact, can be a reasonable one for use in developing an approximate nonlinear asymptotic theory if $\delta\Theta$ is small, i.e., if the OLS estimated parameter is close to θ_0.

From these calculations, we see that the rank of $\chi(\theta_0)$ and the conditioning (or ill-conditioning) of $\mathcal{F}(\theta_0)$ play a significant role in solving OLS inverse problems as well as in any asymptotic standard error formulations based on this linearization. Observe that the error (or noise) $\mathcal{E}$ in the data will in general be amplified as the ill-conditioning of $\mathcal{F}$ increases. We further note that the $N \times p$ sensitivity matrix $\chi(\theta_0)$ is of full rank p if and only if the $p \times p$ Fisher matrix $\mathcal{F}(\theta_0)$ has rank p, or equivalently, is nonsingular. These underlying considerations have motivated a number of efforts (e.g., see [9, 10, 11]) on understanding the conditioning of the Fisher matrix as a function of the number N and longitudinal locations $\{t_j\}_{j=1}^N$ of data points as a key indicator for well-formulated inverse problems and as a tool in optimal design, especially with respect to computation of uncertainty (standard errors, confidence intervals) for parameter estimates.

In view of the considerations above (which are very *local* in nature – both the sensitivity matrix and the Fisher information matrix are taken at the nominal vector θ_0), one should be pessimistic about using these quantities to obtain any *nonlocal* selection methods or criteria for estimation. Indeed, for nonlinear complex systems, it is easy to argue that questions related to some type of global parameter identifiability are not fruitful questions to be pursuing.

3 Subset Selection Algorithm

The focus of our presentation here is how one chooses *a priori* (i.e., *before* any inverse problem calculations are carried out) which parameters can be readily estimated with a typical longitudinal data set. We illustrate an algorithm, developed recently in [28], to select parameter vectors that can be estimated from a given data set using an ordinary least squares inverse problem formulation (similar ideas apply if one is using a relative error statistical model and generalized least squares formulations). Let $q \in \mathbb{R}^{p_0}$ be the parameter vectors being at our disposal for parameter estimation and denote by $q_0 \in \mathbb{R}^{p_0}$ the vector of the corresponding nominal values. Given a number $p < p_0$ of parameters we wish to identify, the algorithm searches all possible choices of p different parameters among the p_0 parameters and selects the one which is identifiable (i.e., the corresponding sensitivity matrix has full rank p) and minimizes a given uncertainty quantification (e.g., by means of asymptotic standard errors). Prior knowledge of a nominal set of values for all parameters along with the observation times for data (but not the values of the observations) will be required for our algorithm. For $p < p_0$ we set

$$\mathcal{S}_p = \left\{ \theta \in \mathbb{R}^p |\ \theta \text{ is a sub-vector of } q \in \mathbb{R}^{p_0} \right\},$$

i.e., $\theta \in \mathcal{S}_p$ is given as $\theta = \left(q_{i_1}, \ldots, q_{i_p}\right)^{\mathsf{T}}$ for some $1 \le i_1 < \cdots < i_p \le p_0$. The corresponding nominal vector is $\theta_0 = \left((q_0)_{i_1}, \ldots, (q_0)_{i_p}\right)^{\mathsf{T}}$.

As we have stated above, to apply the parameter subset selection algorithm we require prior knowledge of nominal variance and nominal parameter values. These nominal values of σ_0 and θ_0 are needed to calculate the sensitivity matrix, the Fisher matrix and the corresponding covariance matrix defined in (12). For our illustration presented in Section 5, we

use the variance and parameter estimates obtained in previous investigations of the models as nominal values. In problems for which no prior estimation has been carried out, one must use knowledge of the observation process error and some knowledge of viable parameter values that might be reasonable with the model under investigation.

The uncertainty quantification we shall use is based on the considerations given in the previous section. Let $\theta \in \mathbb{R}^p$ be given. As an approximation to the covariance matrix of the estimator for θ we take

$$\Sigma(\theta_0) = \sigma_0^2\big(\chi(\theta_0)^{\mathsf{T}}\chi(\theta_0)\big)^{-1} = \mathcal{F}(\theta_0)^{-1}.$$

We introduce the *coefficients of variation* for θ

$$\nu_i(\theta_0) = \frac{\big(\Sigma(\theta_0)_{i,i}\big)^{1/2}}{(\theta_0)_i}, \quad i = 1, \ldots, p, \tag{16}$$

and take as a uncertainty quantification for the estimates of θ the *selection score* given by the Euclidean norm in $\theta \in \mathbb{R}^p$ of $\nu(\theta_0)$, i.e.,

$$\alpha(\theta_0) = |\nu(\theta_0)|,$$

where $\nu(\theta_0) = \big(\nu_1(\theta_0), \ldots, \nu_p(\theta_0)\big)^{\mathsf{T}}$. The components of the vector $\nu(\theta_0)$ are the ratios of each standard error for a parameter to the corresponding nominal parameter value. These ratios are dimensionless numbers warranting comparison even when parameters have considerably different scales and units (e.g., in case of the HIV-model N_T is on the order of 10^1, while k_1 is on the order of 10^{-6}, whereas in case of the CVS-model we have c_ℓ on the order 10^{-2} and $A^{\text{exer}}_{\text{pesk}}$ on the order 10^2). A selection score $\alpha(\theta_0)$ near zero indicates lower uncertainty possibilities in the estimation, while large values of $\alpha(\theta_0)$ suggest that one could expect to find substantial uncertainty in at least some of the components of the estimates in any parameter estimation attempt.

Let $\mathcal{F}_0$ be the Fisher information matrix corresponding to the parameter vectors $q \in \mathbb{R}^{p_0}$ and $\mathcal{F}_p$ the Fisher information matrix corresponding to the parameter vectors in $\theta \in \mathcal{S}_p$. Then $\operatorname{rank}\mathcal{F}_0(q_0) = p_0$ implies that $\operatorname{rank}\mathcal{F}_p(\theta_0) = p$ for any $\theta \in \mathcal{S}_p$, $p < p_0$, i.e., if $\mathcal{F}_0(q_0)$ is non-singular then also all $\mathcal{F}_p(\theta_0)$ are non-singular for all $p < p_0$ and all θ_0 corresponding to a $\theta \in \mathcal{S}_p$. Moreover, if $\operatorname{rank}\mathcal{F}_0(q_0) = p_1$ with $p_1 < p_0$, then $\operatorname{rank}\mathcal{F}(\theta_0) < p$ for all $p_1 < p < p_0$ and all $\theta \in \mathcal{S}_p$.

On the basis of the considerations given above our algorithm proceeds as follows:

Selection Algorithm. *Given $p < p_0$ the algorithm considers all possible choices of indices $i_1, \ldots, i_p$ with $1 \le i_1 < \cdots < i_p \le p_0$ in lexicographical ordering starting with the first choice $(i_1^{(1)}, \ldots, i_p^{(1)}) = (1, \ldots, p)$ and completes the following steps:*

Initializing step: Set $\text{ind}^{\text{sel}} = (1, \ldots, p)$ *and* $\alpha^{\text{sel}} = \infty$.

Step k: For the choice $(i_1^{(k)}, \ldots, i_p^{(k)})$ *compute* $r = \operatorname{rank}\mathcal{F}\big((q_0)_{i_1^{(k)}}, \ldots, (q_0)_{i_p^{(k)}}\big)$.

If $r < p$, go to Step $k+1$.

If $r = p$, compute $\alpha_k = \alpha\big((q_0)_{i_1^{(k)}}, \ldots, (q_0)_{i_p^{(k)}}\big)$.

If $\alpha_k \ge \alpha^{\text{sel}}$, go to Step $k+1$.

If $\alpha_k < \alpha^{\text{sel}}$, set $\text{ind}^{\text{sel}} = (i_1^{(k)}, \ldots, i_p^{(k)})$, $\alpha^{\text{sel}} = \alpha_k$ *and go to Step $k+1$.*

Following the initializing step the algorithm performs $\binom{p_0}{p}$ steps and provides the index vector $\text{ind}^{\text{sel}} = \big(i_1^*, \ldots, i_p^*\big)$ which gives the sub-vector $\theta^* = \big(q_{i_1^*}, \ldots, q_{i_p^*}\big)^{\mathsf{T}}$ such that the

selection score $\alpha\big((q_0)_{i_1^*},\ldots,(q_0)_{i_p^*}\big)$ is minimal among all possible choices of sub-vectors in $\mathcal{S}_p$. If rank $\mathcal{F}_{p_0} = p_0$ then the rank test in Step k can be canceled, of course.

4 Models

In the following we shall illustrate the parameter selection ideas with results obtained by use of the subset selection algorithm described in the previous section for two specific models. These models have a moderate number of parameters to be identified yet are sufficiently complex to make a trial-error approach unfeasible.

4.1 A Mathematical Model for HIV Progression with Treatment Interruption

As our first illustrative example we use one of the many dynamic models for HIV progression found in an extensive literature (e.g., see [1, 2, 3, 4, 8, 20, 26, 50, 52, 58] and the many references therein). For our example model, the dynamics of in-host HIV is described by the interactions between uninfected and infected type 1 target cells (T_1 and T_1^*) (CD4$^+$ T-cells), uninfected and infected type 2 target cells (T_2 and T_2^*) (such as macrophages or memory cells, etc.), infectious free virus V_I, and immune response E (cytotoxic T-lymphocytes CD8$^+$) to the infection. This model, which was developed and studied in [1, 4] and later extended in subsequent efforts (e.g., see [8]), is essentially one suggested in [26], but includes an immune response compartment and dynamics as in [20]. The model equations are given by

$$\begin{aligned}
\dot{T}_1 &= \lambda_1 - d_1 T_1 - \big(1-\bar{\epsilon}_1(t)\big)k_1 V_I T_1,\\
\dot{T}_1^* &= \big(1-\bar{\epsilon}_1(t)\big)k_1 V_I T_1 - \delta T_1^* - m_1 E T_1^*,\\
\dot{T}_2 &= \lambda_2 - d_2 T_2 - (1-f\bar{\epsilon}_1(t))k_2 V_I T_2,\\
\dot{T}_2^* &= \big(1-f\bar{\epsilon}_1(t)\big)k_2 V_I T_2 - \delta T_2^* - m_2 E T_2^*,\\
\dot{V}_I &= \big(1-\bar{\epsilon}_2(t)\big)10^3 N_T \delta (T_1^*+T_2^*) - cV_I\\
&\qquad - \big(1-\bar{\epsilon}_1(t)\big)10^3 k_1 T_1 V_I - (1-f\bar{\epsilon}_1(t))10^3 k_2 T_2 V_I,\\
\dot{E} &= \lambda_E + \frac{b_E\big(T_1^*+T_2^*\big)}{T_1^*+T_2^*+K_b}E - \frac{d_E(T_1^*+T_2^*)}{T_1^*+T_2^*+K_d}E - \delta_E E,
\end{aligned} \tag{17}$$

together with an initial condition vector

$$\big(T_1(0), T_1^*(0), T_2(0), T_2^*(0), V_I(0), E(0)\big)^{\mathsf{T}}.$$

The differences in infection rates and treatment efficacy help create a low, but non-zero, infected cell steady state for T_2^*, which is compatible with the idea that macrophages or memory cells may be an important source of virus after T-cell depletion. The populations of uninfected target cells T_1 and T_2 may have different source rates λ_i and natural death rates d_i. The time-dependent treatment factors $\bar{\epsilon}_1(t) = \epsilon_1 u(t)$ and $\bar{\epsilon}_2(t) = \epsilon_2 u(t)$ represent the effective treatment impact of a reverse transcriptase inhibitor (RTI) (that blocks new infections) and a protease inhibitor (PI) (which causes infected cells to produce non-infectious virus), respectively. The RTI is potentially more effective in type 1 target cells (T_1 and T_1^*) than in type 2 target cells (T_2 and T_2^*), where the efficacy is $f\bar{\epsilon}_1$, with $f \in [0,1]$. The relative effectiveness of RTIs is modeled by ϵ_1 and that of PIs by ϵ_2, while the time-dependent

treatment function $0 \leq u(t) \leq 1$ represents therapy levels, with $u(t) = 0$ for fully off and $u(t) = 1$, for fully on. Although HIV treatment is nearly always administered as combination therapy, the model allows the possibility of mono-therapy, even for a limited period of time, implemented by considering separate treatment functions $u_1(t), u_2(t)$ in the treatment factors.

As in [1, 4], for our numerical investigations we consider a log-transformed and reduced version of the model. This transformation is frequently used in the HIV modeling literature because of the large differences in orders of magnitude in state values in the model and the data and to guarantee non-negative state values as well as because of certain probabilistic considerations (for further discussions see [4]). This results in the nonlinear system of differential equations

$$
\begin{aligned}
\dot{x}_1 &= \frac{10^{-x_1}}{\ln(10)}\Big(\lambda_1 - d_1 10^{x_1} - \big(1-\bar{\varepsilon}_1(t)\big)k_1 10^{x_5}10^{x_1}\Big),\\
\dot{x}_2 &= \frac{10^{-x_2}}{\ln(10)}\Big(\big(1-\bar{\varepsilon}_1(t)\big)k_1 10^{x_5}10^{x_1} - \delta 10^{x_2} - m_1 10^{x_6}10^{x_2}\Big),\\
\dot{x}_3 &= \frac{10^{-x_3}}{\ln(10)}\Big(\lambda_2 - d_2 10^{x_3} - \big(1-f\bar{\varepsilon}_1(t)\big)k_2 10^{x_5}10^{x_3}\Big),\\
\dot{x}_4 &= \frac{10^{-x_4}}{\ln(10)}\Big(\big(1-f\bar{\varepsilon}_1(t)\big)k_2 10^{x_5}10^{x_3} - \delta 10^{x_4} - m_2 10^{x_6}10^{x_4}\Big),\\
\dot{x}_5 &= \frac{10^{-x_5}}{\ln(10)}\Big(\big(1-\bar{\varepsilon}_2(t)\big)10^3 N_T\delta\big(10^{x_2}+10^{x_4}\big) - c10^{x_5}\\
&\qquad - \big(1-\bar{\varepsilon}_1(t)\big)\rho_1 10^3 k_1 10^{x_1}10^{x_5} - \big(1-f\bar{\varepsilon}_1(t)\big)\rho_2 10^3 k_2 10^{x_3}10^{x_5}\Big),\\
\dot{x}_6 &= \frac{10^{-x_6}}{\ln(10)}\Bigg(\lambda_E + \frac{b_E\big(10^{x_2}+10^{x_4}\big)}{10^{x_2}+10^{x_4}+K_b}10^{x_6}\\
&\qquad\qquad - \frac{d_E\big(10^{x_2}+10^{x_4}\big)}{10^{x_2}+10^{x_4}+K_d}10^{x_6} - \delta_E 10^{x_6}\Bigg),
\end{aligned}
\tag{18}
$$

where the changes of variables are defined by

$$T_1 = 10^{x_1},\ T_1^* = 10^{x_2},\ T_2 = 10^{x_3},\ T_2^* = 10^{x_4},\ V_I = 10^{x_5},\ E = 10^{x_6}.$$

The initial conditions for equations (18) are denoted by $x_i(t_0) = x_i^0$, $i = 1, \ldots, 6$. We note that this model has six state variables and the following 22 (in general, unknown) system parameters in the right-hand sides of equations (18)

$$\lambda_1, d_1, \epsilon_1, k_1, \lambda_2, d_2, f, k_2, \delta, m_1, m_2, \ldots \\ \ldots \epsilon_2, N_T, c, \rho_1, \rho_2, \lambda_E, b_E, K_b, d_E, K_d, \delta_E.$$

We will also consider the initial conditions as unknowns and thus we have 28 unknown parameters which we collect in the parameter vector θ,

$$\theta = \big(x_1^0, x_2^0, x_3^0, x_4^0, x_5^0, x_6^0, \lambda_1, d_1, \epsilon_1, k_1, \lambda_2, d_2, f, k_2, \delta, m_1, m_2, \ldots \\ \ldots \epsilon_2, N_T, c, \rho_1, \rho_2, \lambda_E, b_E, K_b, d_E, K_d, \delta_E\big)^{\mathsf{T}}.$$

A list of the parameters in the model equations along with their units is given below in Table 1.

As reported in [1, 4], data to be used with this model in inverse or parameter estimation problems typically consisted of observations for $T_1 + T_1^*$ and V over some extended time

Table 1: Parameters in the equations of the HIV model.

Parameter	Units	Description
λ_1	cells/(ml day)	production rate of type 1 target cells
d_1	1/day	death rate of type 1 target cells
ϵ_1	—	treatment efficacy of type 1 target cells
k_1	ml/(virion day)	infection rate of type 1 target cells
λ_2	cells/(ml day)	production rate of type 2 target cells
d_2	1/day	death rate of type 2 target cells
f	—	reduction of treatment efficacy for type 2 target cells
k_2	ml/(virion day)	infection rate of type 2 target cells
δ	1/day	death rate of infected cells
m_1	ml/(cell day)	immune-induced clearance rate for type 1 target cells
m_2	ml/(cell day)	immune-induced clearance rate for type 2 target cells
ϵ_2	—	treatment efficacy for type 2 target cells
N_T	virions/cell	virions produced per infected cell
c	1/day	natural death rate of viruses
ρ_1	virions/cell	average number of virions infecting a type 1 cell
ρ_2	virions/cell	average number of virions infecting a type 2 cell
λ_E	cells/(ml day)	production rate for immune effectors
b_E	1/day	maximum birth rate for immune effectors
K_b	cells/ml	saturation constant for immune effector birth
d_E	1/day	maximum death rate for immune effectors
K_d	cells/ml	saturation constant for immune effector death
δ_E	1/day	natural death rate for immune effectors

period. For the purpose of this paper we are only using the data for $T_1 + T_1^*$ in case of patient # 4 which we depict in Figure 1 together with the corresponding model output for the parameters identified in [1, 4]. Thus our observations are

$$f(t_i;\theta_0) = \log_{10}\left(10^{x_1(t_i;\theta_0)} + 10^{x_2(t_i;\theta_0)}\right), \tag{19}$$

where the nominal parameter vector θ_0 is given at the beginning of Subsection 5.1.

While the inverse problem we are considering in this paper for the HIV-model is relatively "small" compared to many of those found in the literature, it is still represents a nontrivial estimation challenge and is more than sufficient to illustrate the ideas and methodology we discuss in this presentation. Other difficult aspects (censored data requiring use of the Expectation Maximization algorithm as well as use of residual plots in attempts to validate the correctness of choice of corresponding statistical models introduced and discussed in Section 2) of such inverse problems are discussed in the review chapter [7] and will not be pursued here.

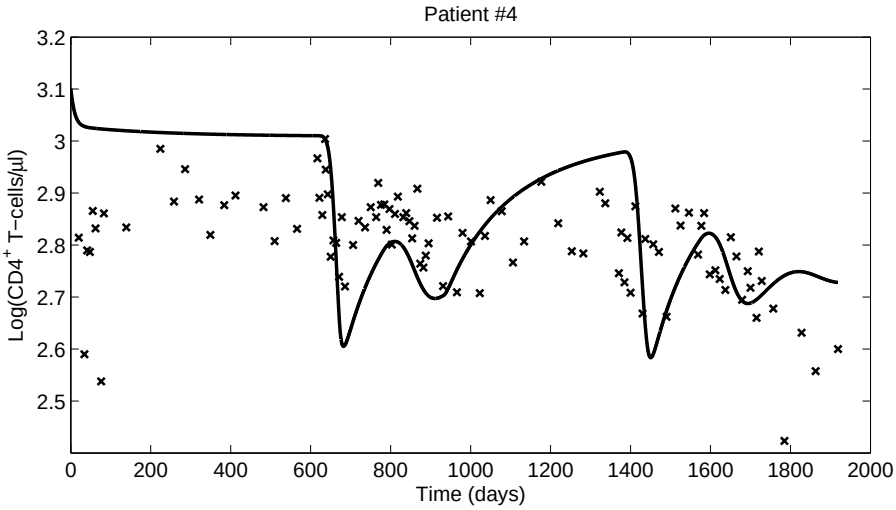


Figure 1: Log-scaled data $\{y_j\}$ for CD4$^+$ T-cells of Patient #4 (crosses), and model output $z(t)$ (solid curve) evaluated at the parameter estimates obtained in [1, 4].

4.2 A model for the reaction of the cardiovascular system to an ergometric workload

As a second example to illustrate our methods we chose a model for cardiovascular function. The model was developed in order to describe the reaction of the cardiovascular system to a constant ergometric workload of moderate size. The building block of the model are the left ventricle, the arterial and venous systemic compartments representing the systemic circuit as well as the right ventricle, arterial and venous pulmonary compartments representing the pulmonary circuit. The model is non-pulsatile and includes the baroreceptor loop as the essential control loop in the situation to be modeled. The feedback control which steers the system from the equilibrium corresponding to rest to the equilibrium corresponding to the imposed workload is obtained by solving a linear quadratic regulator problem. Furthermore the model includes a sub-model describing the so called autoregulation process, i.e., the control of local blood flow in response to local metabolic demands. The model equations are given as (for details see [45], [16, Chapter 1]):

$$
\begin{aligned}
\dot{P_{\mathrm{as}}} &= \frac{1}{c_{\mathrm{as}}}\Big(Q_\ell - \frac{1}{R_{\mathrm{s}}}(P_{\mathrm{as}} - P_{\mathrm{vs}})\Big),\\
\dot{P_{\mathrm{vs}}} &= \frac{1}{c_{\mathrm{vs}}}\Big(\frac{1}{R_{\mathrm{s}}}(P_{\mathrm{as}} - P_{\mathrm{vs}}) - Q_{\mathrm{r}}\Big),\\
\dot{P_{\mathrm{ap}}} &= \frac{1}{c_{\mathrm{ap}}}\Big(Q_{\mathrm{r}} - \frac{1}{R_{\mathrm{p}}}(P_{\mathrm{ap}} - P_{\mathrm{vp}}(P_{\mathrm{as}}, P_{\mathrm{vs}}, P_{\mathrm{ap}}))\Big),\\
\dot{S}_\ell &= \sigma_\ell,\\
\dot{\sigma}_\ell &= -\alpha_\ell S_\ell - \gamma_\ell \sigma_\ell + \beta_\ell H,\\
\dot{S}_{\mathrm{r}} &= \sigma_{\mathrm{r}},\\
\dot{\sigma}_{\mathrm{r}} &= -\alpha_{\mathrm{r}} S_{\mathrm{r}} - \gamma_{\mathrm{r}} \sigma_{\mathrm{r}} + \beta_{\mathrm{r}} H,\\
\dot{R}_{\mathrm{s}} &= \frac{1}{K}\Big(A_{\mathrm{pesk}}\Big(\frac{P_{\mathrm{as}} - P_{\mathrm{vs}}}{R_{\mathrm{s}}} C_{\mathrm{a,O_2}} - M_0 - \rho W\Big) - (P_{\mathrm{as}} - P_{\mathrm{vs}})\Big),\\
\dot{H} &= u(t)
\end{aligned}
\tag{20}
$$

with

$$
\begin{aligned}
P_{\mathrm{vp}} &= P_{\mathrm{vp}}(P_{\mathrm{as}}, P_{\mathrm{vs}}, P_{\mathrm{ap}}) := \frac{1}{c_{\mathrm{vp}}}\big(V_{\mathrm{tot}} - c_{\mathrm{as}} P_{\mathrm{as}} - c_{\mathrm{vs}} P_{\mathrm{vs}} - c_{\mathrm{ap}} P_{\mathrm{ap}}\big),\\
Q_\ell &= H \frac{c_\ell P_{\mathrm{vp}}(P_{\mathrm{as}}, P_{\mathrm{vs}}, P_{\mathrm{ap}}) a_\ell(H) S_\ell}{a_\ell(H) P_{\mathrm{as}} + k_\ell(H) S_\ell},\\
Q_{\mathrm{r}} &= H \frac{c_{\mathrm{r}} P_{\mathrm{vs}} a_{\mathrm{r}}(H) S_{\mathrm{r}}}{a_{\mathrm{r}}(H) P_{\mathrm{ap}} + k_{\mathrm{r}}(H) S_{\mathrm{r}}},
\end{aligned}
\tag{21}
$$

where

$$
\begin{aligned}
k_\ell(H) &= e^{-(c_\ell R_\ell)^{-1} t_d(H)} \quad \text{and} \quad a_\ell(H) = 1 - k_\ell(H),\\
k_{\mathrm{r}}(H) &= e^{-(c_{\mathrm{r}} R_{\mathrm{r}})^{-1} t_d(H)} \quad \text{and} \quad a_{\mathrm{r}}(H) = 1 - k_{\mathrm{r}}(H).
\end{aligned}
\tag{22}
$$

For the duration t_d of the diastole we use Bazett's formula (duration of the systole $= \kappa/H^{1/2}$) which implies

$$
t_d = t_d(H) = \frac{1}{H^{1/2}}\Big(\frac{1}{H^{1/2}} - \kappa\Big). \tag{23}
$$

Introducing $x = \big(P_{\mathrm{as}}, P_{\mathrm{vs}}, P_{\mathrm{ap}}, S_\ell, \sigma_\ell, S_{\mathrm{r}}, \sigma_{\mathrm{r}}, R_{\mathrm{s}}, H\big)^{\mathsf{T}} \in \mathbb{R}^9$ system (20) can be written as

$$
\dot{x}(t) = f(x(t), W, \theta, u(t)),
$$

where $W = W^{\mathrm{rest}} = 0$ (Watt) for $t <= 0$ and $W = W^{\mathrm{exer}} = 75$ (Watt) for $t > 0$. Moreover, θ is the parameter vector of the system. We distinguish two values for each of the parameters R_{p} and A_{pesk}, one for the resting situation and one for the exercise situation. Consequently we have the parameters $R_{\mathrm{p}}^{\mathrm{rest}}$, $A_{\mathrm{pesk}}^{\mathrm{rest}}$ and $R_{\mathrm{p}}^{\mathrm{exer}}$ and $A_{\mathrm{pesk}}^{\mathrm{exer}}$ instead of R_{p} and A_{pesk}. The initial value for system (20) is the equilibrium x^{rest}, which is computed from $f(x^{\mathrm{rest}}, W^{\mathrm{rest}}, \theta, 0) = 0$. Analogously x^{exer} is the equilibrium corresponding to the constant workload W^{exer} (satisfying $f(x^{\mathrm{exer}}, W^{\mathrm{exer}}, \theta, 0) = 0$).

Let $B = \big(0, \ldots, 0, 1\big)^{\mathsf{T}} \in \mathbb{R}^9$, $C = \big(1, 0, \ldots, 0\big) \in \mathbb{R}^{1\times 9}$ and $A(\theta) = \frac{\partial f(x, W^{\mathrm{exer}}, \theta, 0)}{\partial x}\big|_{x = x^{\mathrm{exer}}}$, the Jacobian of the right-hand side of system (20) at the equilibrium x^{exer}. The control $u(t)$ is obtained as the solution of the linear-quadratic regulator problem for the linear system

$$
\dot{\xi}(t) = A(\theta)\xi(t) + Bu(t), \quad \xi(0) = x^{\mathrm{rest}} - x^{\mathrm{exer}}, \tag{24}
$$

where we have set $\xi(t) = x(t) - x^{\text{exer}}$, and the quadratic cost functional

$$J(u) = \int_0^\infty \big(q_{\text{as}}^2 (P_{\text{as}}(t) - P_{\text{as}}^{\text{exer}})^2 + u(t)^2\big)\, dt. \tag{25}$$

This functional reflects the fact that the baroreceptor loop, which is the basic control loop for the situation we are considering here, generates the control using the arterial systemic pressure $P_{\text{as}}(t)$.

According to the theory of the linear-quadratic control problem $u(t)$ is given by

$$u(t) = -B^{\mathsf{T}} X \xi(t) = -B^{\mathsf{T}} X (x(t) - x^{\text{exer}}), \quad t \geq 0, \tag{26}$$

where $X = X(\theta)$ is the solution of the Riccati matrix equation $XA(\theta) + A(\theta)^{\mathsf{T}} X - XBB^{\mathsf{T}} X + C^{\mathsf{T}} C = 0$. The feedback control (26) is also a stabilizing control for system (20), i.e., we have $\lim_{t\to\infty} x(t) = x^{\text{exer}}$ provided $\|x(0) - x^{\text{exer}}\|_2$ is sufficiently small.

Table 2: Parameters of the CVS-model.

Parameter	Units	Description
c_{as}	liter/mmHg	compliance of the arterial systemic compartment
c_{vs}	liter/mmHg	compliance of the venous systemic compartment
c_{ap}	liter/mmHg	compliance of the arterial pulmonary compartment
c_{vp}	liter/mmHg	compliance of the venous pulmonary compartment
c_ℓ	liter/mmHg	compliance of the relaxed left ventricle
c_{r}	liter/mmHg	compliance of the relaxed right ventricle
V_{tot}	liter	total blood volume
R_{p}	mmHg min/liter	resistance in the peripheral region of the pulmonary circuit
R_ℓ	mmHg min/liter	inflow resistance of the left ventricle
R_{r}	mmHg min/liter	inflow resistance of the right ventricle
κ	$\text{min}^{1/2}$	coefficient in Bazett's formula (see (23))
$C_{\text{a,O}_2}$	1	O_2-concentration in arterial systemic blood
K	liter	constant in the formula for the biochemical energy flow, $M_b = -K\, dC_{\text{v,O}_2}/dt$
A_{pesk}	mmHg min/liter	constant in the formula relating peripheral systemic resistance and venous O_2 concentration ($R_{\text{s}} = A_{\text{pesk}} C_{\text{v,O}_2}$)
M_0	liter/min	metabolic rate in the systemic tissue region corresponding to zero workload
ρ	liter/(min Watt)	coefficient of W in the differential equation for R_{s}
q_{as}	$\text{min}^{-2}(\text{mmHg})^{-1}$	weighting factor for P_{as} in the cost functional (25)
α_ℓ	min^{-2}	coefficient of S_ℓ in the differential equation for S_ℓ
α_{r}	min^{-2}	coefficient of S_r in the differential equation for S_{r}
β_ℓ	mmHg/min	coefficient of H in the differential equation for S_ℓ
β_{r}	mmHg/min	coefficient of H in the differential equation for S_{r}
γ_ℓ	min^{-1}	coefficient of $\dot{S}_\ell$ in the differential equation for S_ℓ
γ_{r}	min^{-1}	coefficient of $\dot{S}_{\text{r}}$ in the differential equation for S_{r}

The parameter vector of the system is

$$\begin{aligned} q = \big(&c_\ell, c_{\text{r}}, c_{\text{as}}, c_{\text{vs}}, c_{\text{ap}}, c_{\text{vp}}, R_\ell, R_{\text{r}}, \alpha_\ell, \alpha_{\text{r}}, \beta_\ell, \beta_{\text{r}}, \gamma_\ell, \dots \\ &\dots \gamma_{\text{r}}, K, \kappa, M_0, \rho, C_{\text{a,O}_2}, q_{\text{as}}, V_{\text{tot}}, R_{\text{p}}^{\text{rest}}, A_{\text{pesk}}^{\text{rest}}, R_{\text{p}}^{\text{exer}}, A_{\text{pesk}}^{\text{exer}}\big)^{\mathsf{T}} \in \mathbb{R}^{25}. \end{aligned} \tag{27}$$

Tables 2 and 3 contain the descriptions and units for the parameters q and the state variables x, respectively, of the system.

Table 3: The state variables and other variables of the CVS-model.

Variable	Unit	Description
$P_{\rm as}$	mmHg	pressure in the arterial systemic compartment
$P_{\rm vs}$	mmHg	pressure in the venous systemic compartment
$P_{\rm ap}$	mmHg	pressure in the arterial pulmonary compartment
$P_{\rm vp}$	mmHg	pressure in the venous pulmonary compartment
S_ℓ	mmHg	contractility of the left ventricle
σ_ℓ	mmHg/min	time derivative of S_ℓ
$S_{\rm r}$	mmHg	contractility of the right ventricle
$\sigma_{\rm r}$	mmHg/min	time derivative of $S_{\rm r}$
$R_{\rm s}$	mmHg min/liter	peripheral resistance in the systemic circuit
H	$\min^{-1}$	heart rate
Q_ℓ	liter/min	cardiac output of the left ventricle
$Q_{\rm r}$	liter/min	cardiac output of the right ventricle
W	Watt	workload imposed on the test person

5 Results and Discussion

5.1 The HIV-model

As the nominal parameter vector we take the estimates obtained in [1, 4] for Patient # 4. More precisely, we assume that the error variance is $\sigma_0^2 = 0.11$, and that the nominal parameter values (for description and units see Table 1) are given as:

$$
\begin{aligned}
x_1^0 &= \log_{10}(1.202\text{e+}3), & x_2^0 &= \log_{10}(6.165\text{e+}1), & x_3^0 &= \log_{10}(1.755\text{e+}1),\\
x_4^0 &= \log_{10}(6.096\text{e-}1), & x_5^0 &= \log_{10}(9.964\text{e+}5), & x_6^0 &= \log_{10}(1.883\text{e-}1),\\
\lambda_1 &= 4.633, & d_1 &= 4.533\times 10^{-3}, & \epsilon_1 &= 6.017\times 10^{-1},\\
k_1 &= 1.976\times 10^{-6}, & \lambda_2 &= 1.001\times 10^{-1}, & d_2 &= 2.211\times 10^{-2},\\
f &= 5.3915\times 10^{-1}, & k_2 &= 5.529\times 10^{-4}, & \delta &= 1.865\times 10^{-1},\\
m_1 &= 2.439\times 10^{-2}, & m_2 &= 1.3099\times 10^{-2}, & \epsilon_2 &= 5.043\times 10^{-1},\\
N_T &= 1.904\times 10^{1}, & c &= 1.936\times 10^{1}, & \rho_1 &= 1.000,\\
\rho_2 &= 1.000, & \lambda_E &= 9.909\times 10^{-3}, & b_E &= 9.785\times 10^{-2},\\
K_b &= 3.909\times 10^{-1}, & d_E &= 1.021\times 10^{-1}, & K_d &= 8.379\times 10^{-1},\\
\delta_E &= 7.030\times 10^{-2}.
\end{aligned}
$$

In Figure 1 above we depicted the log-scaled longitudinal observations $\{y_i\}$ on the number of CD4$^+$ T-cells and the model output $z(t_j) = f(t_j, \theta_0)$, $j = 1, \ldots, N$, evaluated at the nominal parameter vector and given in (19).

It is assumed that the following parameters are always fixed at the values given above:

$$x_3^0,\ x_4^0,\ x_6^0,\ \rho_1,\ \text{and}\ \rho_2.$$

In other words, the parameters to be selected for estimation will always constitute a sub-vector of

$$
\begin{aligned}
q = \big(&x_1^0, x_2^0, x_5^0, \lambda_1, d_1, \epsilon_1, k_1, \lambda_2, d_2, f, k_2, \delta, \ldots\\
&\ldots m_1, m_2, \epsilon_2, N_T, c, \lambda_E, b_E, K_b, d_E, K_d, \delta_E\big) \in \mathbb{R}^{23}. \quad (28)
\end{aligned}
$$

Let $\mathcal{F}_{23}$ denote the Fisher information matrix of system (18) for the 23 parameters of q as given in (28) at their nominal values q_0 as given above. Then we have

$$\operatorname{cond} \mathcal{F}_{23}(q_0) = 1.712 \times 10^{24},$$

i.e., $\mathcal{F}_{23}(q_0)$ is non-singular, but ill-conditioned. Since $\mathcal{F}_{23}(q_0)$ is non-singular, the Fisher information matrix for any sub-vector θ of q at the corresponding nominal parameter values is also non-singular. Consequently the regularity check for the Fisher information matrix in the subset selection algorithm can be deleted in case of the HIV-model.

Table 4: The top five parameter vectors obtained with subset selection algorithm for $p = 11$. For each parameter vector θ the condition number $\kappa(\mathcal{F}(\theta_0))$ of the Fisher information matrix and the selection score $\alpha(\theta_0)$ are displayed. The next two lines show $\kappa(\mathcal{F}(\theta_0))$ and $\alpha(\theta_0)$ for the sub-vector $\theta \in \mathbb{R}^9$ which is common to the top five parameter vectors and for the optimal parameter vector in $\mathbb{R}^9$. The last line presents the sub-vector in $\mathbb{R}^{11}$ with the largest selection score.

Parameter vector θ	$\kappa(\mathcal{F}(\theta_0))$	$\alpha(\theta_0)$
$(x_1^0,\ x_5^0,\ \lambda_1,\ d_1,\ \epsilon_1,\ \lambda_2,\ d_2,\ k_2,\ \delta,\ \epsilon_2,\ N_T)$	9.841×10^{10}	2.881×10^{1}
$(x_1^0,\ x_5^0,\ \lambda_1,\ d_1,\ \epsilon_1,\ \lambda_2,\ d_2,\ k_2,\ \delta,\ \epsilon_2,\ c)$	9.845×10^{10}	2.883×10^{1}
$(x_1^0,\ x_5^0,\ \lambda_1,\ d_1,\ \epsilon_1,\ k_1,\ \lambda_2,\ d_2,\ k_2,\ \delta,\ \epsilon_2)$	4.388×10^{16}	2.896×10^{1}
$(x_2^0,\ x_5^0,\ \lambda_1,\ d_1,\ \epsilon_1,\ \lambda_2,\ d_2,\ k_2,\ \delta,\ \epsilon_2,\ N_T)$	9.235×10^{10}	2.904×10^{1}
$(x_2^0,\ x_5^0,\ \lambda_1,\ d_1,\ \epsilon_1,\ \lambda_2,\ d_2,\ k_2,\ \delta,\ \epsilon_2,\ c)$	9.241×10^{10}	2.906×10^{1}
$(x_5^0,\ \lambda_1,\ d_1,\ \epsilon_1,\ \lambda_2,\ d_2,\ k_2,\ \delta,\ \epsilon_2)$	9.083×10^{10}	2.193×10^{1}
$(x_1^0,\ x_5^0,\ \lambda_1,\ d_1,\ k_1,\ d_2,\ k_2,\ \delta,\ \epsilon_2)$	4.335×10^{15}	1.050×10^{1}
$(d_2,\ k_2,\ \delta,\ m_2,\ N_T,\ \lambda_E,\ b_E,\ K_b,\ d_E,\ K_d,\ \delta_E)$	7.247×10^{17}	2.340×10^{5}

In [1, 4], the authors estimate the parameter vector

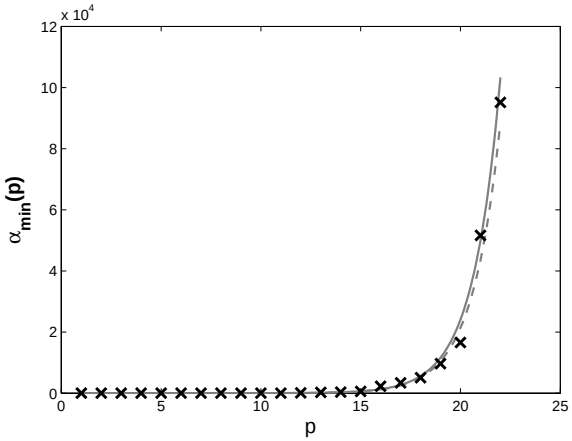
$$\theta = \left(x_1^0, x_2^0, x_5^0, \lambda_1, d_1, \epsilon_1, k_1, \epsilon_2, N_T, c, b_E\right)^{\mathsf{T}} \in \mathbb{R}^{11}. \tag{29}$$

The selection score for this parameter vector is $\alpha(\theta_0) = 4.611 \times 10^3$. In Table 4 we display, for the five selections of sub-vector $\theta \in \mathbb{R}^{11}$ of q with the minimal selection scores, the condition numbers of the corresponding Fisher information matrices and the selection scores. In addition we also show the selection $\theta \in \mathbb{R}^{11}$ with the maximal selection score. As we can see, the selection score values range from 2.881×10^1 to 2.340×10^5 for the $\binom{23}{11} = 1\,352\,078$ different parameter vectors in $\mathbb{R}^{11}$ which can be selected from the 23 parameters in (28).

As we also can see from Table 4 that the selection algorithm chooses most of the parameters in the vector (29). For instance, the sub-vector $(x_5^0, \lambda_1, d_1, \epsilon_1, \epsilon_2)$ of (29) appears in every one of the top five parameter vectors displayed in Table 4. In fact the top five parameter vectors have the sub-vector $\left(x_5^0, \lambda_1, d_1, \epsilon_1, \lambda_2, d_2, k_2, \delta, \epsilon_2\right) \in \mathbb{R}^9$ in common and differ only by one or two of the parameters x_1^0, x_2^0, k_1, c, and N_T. Use of the subset selection algorithm discussed here (had it been available) might have proved valuable in the efforts reported in [1, 4].

In Figure 2(a) we depict the minimal selection score as a function of the number of parameters. Table 5 contains the values of the corresponding selection scores. Figure 2(b) is a semilog plot of Figure 2(a), i.e., it displays the logarithm of the selection score as a function of the number of parameters. Figure 2, (b), suggests that parameter vectors with more than thirteen parameters might be expected to have large uncertainty when estimated from

(a)



(b)

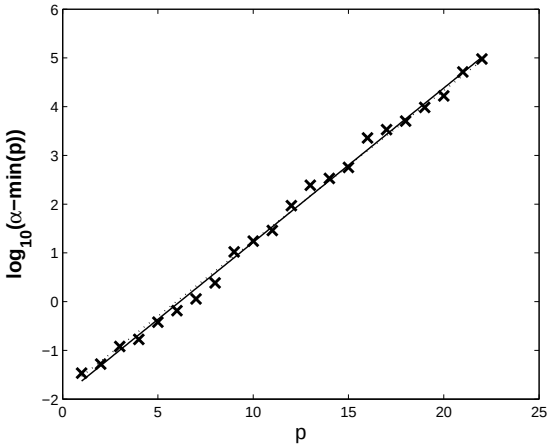


Figure 2: (a) Minimal selection scores (crosses) and exponential approximations (30) (grey solid line) respectively (31) (grey dashed line) versus the number of parameters p. (b) Logarithm of minimal selection scores (crosses) and regression lines corresponding to (30) (gray solid line) respectively to (31) (grey dashed line) versus number of parameters p.

Table 5: Minimal selection scores $\alpha_{\min}(p)$ for sub-vectors in $\mathbb{R}^p$ of (28), $p = 1, 2, \ldots, 22$.

p	1	2	3	4	5	6	7	8
$\alpha_{\min}(p)$	0.0340	0.0523	0.1203	0.1679	9.3796	0.6511	1.1375	2.4166
p	9	10	11	12	13	14	15	16
$\alpha_{\min}(p)$	10.503	17.482	28.81	92.91	243.34	336.77	566.86	2 274.3
p	17	18	19	20	21	22		
$\alpha_{\min}(p)$	3 372.4	5 047.9	9 664.4	16 585	51 631	95 128		

observations, because the minimal selection score is already larger than 100. Figure 2(b) also depicts the regression line, which fits the logarithm of the selection score. From this linear regression we conclude the selection score $\alpha_{\min}(p)$ grows exponentially with the number p of parameters to be estimated. More precisely, we find

$$\alpha_{\min}(p) \approx 0.01133 e^{0.728p}, \quad p = 1, \ldots, 22. \tag{30}$$

Computing the minimal selection scores for $p = 1, \ldots, 22$ requires one to consider all possible choices for sub-vectors of (28) in $\mathbb{R}^p$, $p = 1, \ldots, 22$, i.e., to consider $\sum_{i=1}^{22} \binom{23}{i} = 8\,388\,605$ cases. If we determine the regression line only using the minimal selection scores for $p = 1, 2, 3, 20, 21, 22$ we obtain

$$\alpha_{\min}(p) \approx 0.01441 e^{0.710p}, \quad p = 1, \ldots, 22. \tag{31}$$

Table 6: Time for computing $\theta \in \mathbb{R}^p$ with the minimal selection score on a laptop computer, once the 23×23 Fisher information matrix for the HIV-model has been computed.

p	1	2	3	4	5	6	7	8
time (sec)	0.054	0.015	0.075	0.472	1.66	5.52	14.05	29.81
p	9	10	11	12	13	14	15	16
time (sec)	53.49	80.02	100.4	105.8	96.74	74.01	47.15	24.09
p	17	18	19	20	21	22		
time (sec)	10.69	3.64	1.02	0.211	0.034	0.0041		

Computing the minimal selection scores needed for (31) requires to consider only $\binom{23}{1} + \binom{23}{2} + \binom{23}{3} + \binom{23}{20} + \binom{23}{21} + \binom{23}{22} = 2\big(\binom{23}{1} + \binom{23}{2} + \binom{23}{3}\big) = 4\,094$ cases. In Figures 2(a), we show the curves given by (30) and (31), whereas in Figures 2(b), also the corresponding regression lines are depicted. Table 6 shows the time it takes on a laptop computer with an Intel© Core™2 Duo processor using a MATLAB programm to compute $\alpha_{\min}(p)$, $p = 1, \ldots, 22$, once the 23×23 Fisher information matrix for the nominal parameter vector q_0 has been computed.

Figure 3 is the same as Figure 2(b), but for $p = 1, \ldots, 11$. The curves (30) respectively (31) can be used to determine p such that the selection score $\alpha_{\min}(p)$ is smaller than a given upper bound. From Figure 3 we can see that in order to have $\alpha_{\min}(p) < 5$ we should choose $p \leq 8$, which is correct according to each of the two curves (see Table 5). In order to have $\alpha_{\min} < 10$ the curves suggest $p \leq 9$, which is not quite correct, because according to

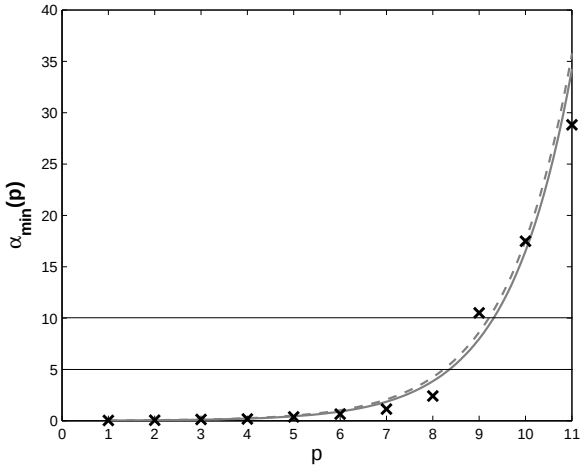


Figure 3: Minimal selection scores (crosses) and and exponential approximation (30) (grey solid line) for $p = 1, \ldots, 11$.

Table 7: The top 5 parameter vectors $\theta \in \mathbb{R}^{11}$ selected according the the criterion of minimal condition number for the Fisher information matrix.

Parameter vector θ	$\kappa(\mathcal{F}(\theta_0))$	$\alpha(\theta_0)$
$(x_1^0,\ x_2^0,\ \lambda_1,\ \epsilon_1,\ f,\ m_1,\ m_2,\ N_T,\ \lambda_E,\ d_E,\ \delta_E)$	1.735×10^6	1.908×10^3
$(x_1^0,\ x_2^0,\ \lambda_1,\ \epsilon_1,\ f,\ m_1,\ m_2,\ N_T,\ \lambda_E,\ b_E,\ \delta_E)$	1.738×10^6	1.897×10^3
$(x_1^0,\ x_2^0,\ \lambda_1,\ \epsilon_1,\ f,\ m_1,\ m_2,\ c,\ \lambda_E,\ d_E,\ \delta_E)$	1.744×10^6	1.908×10^3
$(x_1^0,\ x_2^0,\ \lambda_1,\ \epsilon_1,\ f,\ m_1,\ m_2,\ c,\ \lambda_E,\ b_E,\ \delta_E)$	1.747×10^6	1.896×10^3
$(x_2^0,\ x_5^0,\ \lambda_1,\ \epsilon_1,\ f,\ m_1,\ m_2,\ c,\ \lambda_E,\ d_E,\ \delta_E)$	1.788×10^6	1.910×10^3

Table 5 we have $\alpha_{\min}(9) = 10.5$, so that we should choose $p \leq 8$. In Figure 4 we graph (in logarithmic scales) the condition number $\kappa(\mathcal{F}_p(\theta_0))$ of the corresponding Fisher information matrix versus the smallest selection score $\alpha_{\min}(p)$ for $p = 1, \ldots, 22$. It is clear from Figure 4 that the condition numbers for the Fisher information matrix corresponding to the selected parameter vector $\theta \in \mathbb{R}^p$ does not show a monotone behavior with respect to p. We could also determine the selection of parameters according to the criterion of minimal condition number for the corresponding Fisher information matrix. In Table 7 we present the best 5 selections $\theta \in \mathbb{R}^{11}$ according to this criterion together with the condition numbers of the Fisher information matrix and the corresponding selection scores.

In Table 8 we examine the effect that removing parameters from an estimation has in uncertainty quantification. The coefficient of variation (CV) is shown for each parameter (see (16)). Five cases are considered:

(i) The parameter vector

$$\theta^{(18)} = \left(x_2^0, x_5^0, \lambda_1, d_1, \epsilon_1, k_1, \lambda_2, d_2, f, k_2, m_1, \epsilon_2, c, \lambda_E, d_E, K_d, \delta_E\right)^{\mathsf{T}},$$

which is the sub-vector in $\mathbb{R}^{18}$ with the minimal selection score.

(ii) The parameter vector

$$\theta^{(5,1)} = \left(x_5^0, \lambda_1, d_1, k_1, \epsilon_2\right)^{\mathsf{T}},$$

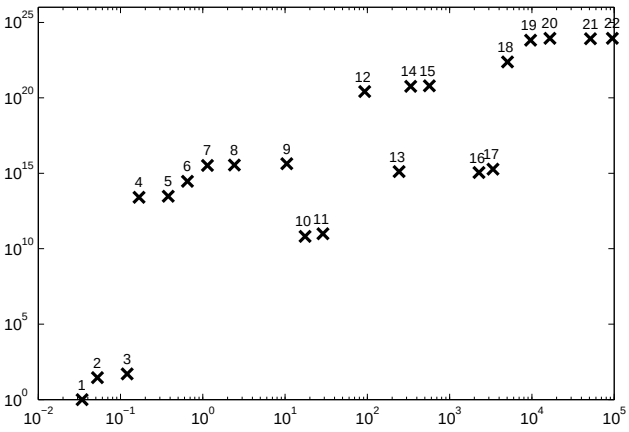


Figure 4: Condition number $\kappa(\mathcal{F}_p(\theta_0))$ versus minimal selection score $\alpha_{\min}(p)$ for the HIV-model, where $\theta \in \mathbb{R}^p$, is the sub-vector of (28) with the minimal selection score, $p = 1, \ldots, 22$. Both axes are in logarithmic scale.

which is the sub-vector of $\theta^{(18)}$ in $\mathbb{R}^5$ with the minimal selection score.

(iii) The parameter vector

$$\theta^{(5,2)} = \left(x_1^0, \lambda_1, k_1, \delta, \epsilon_2\right)^{\mathsf{T}},$$

which is the sub-vector in $\mathbb{R}^5$ of q as given by (28) with the minimal selection score.

(iv) The parameter vector

$$\theta^{(5,3)} = \left(\epsilon_1, \lambda_2, m_1, \epsilon_2, \lambda_E\right)^{\mathsf{T}},$$

which is the sub-vector of $\theta^{(18)}$ in $\mathbb{R}^5$ with the maximal selection score.

(v) The parameter vector

$$\theta^{(5,4)} = \left(m_1, b_E, K_b, d_E, K_d\right)^{\mathsf{T}},$$

which is the sub-vector in $\mathbb{R}^5$ of q as given by (28) with the maximal selection score.

We see that here are substantial improvements in uncertainty quantification when considering $\theta^{(5,1)}$ instead of $\theta^{(18)}$. However, just taking a considerably lower dimensional sub-vector of $\theta^{(18)}$ in $\mathbb{R}^5$ does not lead necessarily to a drastic improvement of the estimate.

Table 8: Coefficient of variation (CV) of the parameter vectors for the HIV-model as listed above

	CV for				
Parameter	$\theta^{(18)}$	$\theta^{(5,1)}$	$\theta^{(5,2)}$	$\theta^{(5,3)}$	$\theta^{(5,4)}$
x_1^0	—	—	4.09×10^{-2}	—	—
x_2^0	3.80	—	—	—	—
x_5^0	1.58×10^{1}	3.43×10^{-1}	—	—	—
λ_1	8.19×10^{-1}	3.56×10^{-1}	1.13×10^{-1}	—	—
d_1	9.39×10^{-1}	3.94×10^{-1}	—	—	—
ϵ_1	1.26×10^{2}	—	—	8.49	—
k_1	7.67×10^{2}	8.17×10^{-2}	9.57×10^{-2}	—	—
λ_2	4.74×10^{1}	—	—	9.99	—
d_2	4.62×10^{1}	—	—	—	—
f	2.51×10^{2}	—	—	—	—
k_2	7.53×10^{2}	—	—	—	—
δ	—	—	3.29×10^{-1}	—	—
m_1	2.29×10^{3}	—	—	9.85×10^{2}	2.10×10^{1}
ϵ_2	1.63×10^{2}	1.11×10^{-1}	9.33×10^{-2}	6.34	—
c	7.74×10^{2}	—	—	—	—
λ_E	2.18×10^{3}	—	—	9.83×10^{2}	—
b_E	—	—	—	—	1.22×10^{4}
K_b	2.55×10^{3}	—	—	—	4.62×10^{3}
d_E	4.56×10^{2}	—	—	—	1.16×10^{4}
K_d	1.98×10^{3}	—	—	—	4.40×10^{3}
δ_E	1.72×10^{3}	—	—	—	—
$\alpha(\cdot)$	5.05×10^{3}	6.47×10^{-1}	3.75×10^{-1}	1.39×10^{3}	1.80×10^{4}

5.2 The CVS-model

For this model we take as the nominal parameters the following estimates obtained in [45] using data obtained at bicycle ergometer tests (for a description and units see Table 2):

$$\begin{aligned}
c_\ell &= 0.02305, & c_{\mathrm{r}} &= 0.04413, & c_{\mathrm{as}} &= 0.01016,\\
c_{\mathrm{vs}} &= 0.6500, & c_{\mathrm{ap}} &= 0.03608, & c_{\mathrm{vp}} &= 0.1408,\\
R_\ell &= 0.2671, & R_{\mathrm{r}} &= 0.04150, & \alpha_\ell &= 30.5587,\\
\alpha_{\mathrm{r}} &= 28.6785, & \beta_\ell &= 25.0652, & \beta_{\mathrm{r}} &= 1.4132,\\
\gamma_\ell &= -1.6744, & \gamma_{\mathrm{r}} &= -1.8607, & K &= 16.0376,\\
\kappa &= 0.05164, & M_0 &= 0.35, & \rho &= 0.011,\\
C_{\mathrm{a,O_2}} &= 0.2, & q_{\mathrm{as}} &= 163.047, & V_{\mathrm{tot}} &= 5.0582,\\
R_{\mathrm{p}}^{\mathrm{rest}} &= 1.5446, & A_{\mathrm{pesk}}^{\mathrm{rest}} &= 177.682, & R_{\mathrm{p}}^{\mathrm{exer}} &= 0.3165,\\
& & A_{\mathrm{pesk}}^{\mathrm{exer}} &= 254.325. & &
\end{aligned}$$

For the estimates given above measurements for P_{as}, H and Q_ℓ were available. The sampling times t_j for the measurements of P_{as} and H were uniformly distributed on the time interval from 0 to 10 minutes with $t_{j+1} - t_j = 2$ seconds, i.e., 601 measurements for P_{as} and H. The measurements for Q_ℓ were obtained by Doppler echo-cardiography and consequently were much less frequent (only 20 measurements) and also irregularly distributed. In the following we shall consider P_{as} as the only measured output of the system, i.e., we have $f(t;\theta) = P_{\mathrm{as}}(t;\theta)$. The variance of the measurement errors was roughly estimated to be $\sigma_0^2 = 10$.

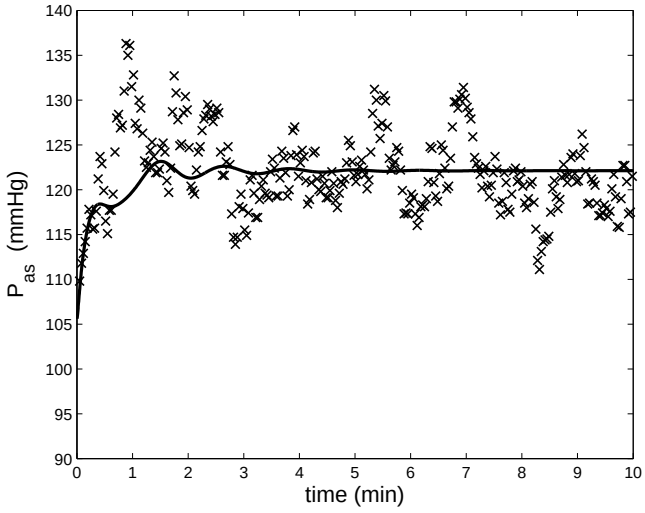


Figure 5: The P_{as}-component of the solution of system (20) with the nominal parameter values and the P_{as}-measurements.

The equilibria x^{rest} and x^{exer} are given in Table 9 (for units see Table 2).

As for the HIV-model also in case of the CVS-model the Fisher information matrix $\mathcal{F}_{25}(q_0)$ at the nominal values of the parameters (27) is non-singular, but highly ill-conditioned:

$$\operatorname{cond} \mathcal{F}_{25}(q_0) = 2.5755 \times 10^{31}.$$

Table 9: The equilibria x^{rest} and x^{exer} for the CVS-model.

variable	P_{as}	P_{vs}	P_{ap}	P_{vp}	S_ℓ	σ_ℓ	S_{r}	σ_{r}	R_{s}	H
x^{rest}	105.595	4.277	12.474	5.367	64.675	0	3.886	0	22.020	78.85
x^{exer}	122.115	3.595	10.441	7.844	88.092	0	5.293	0	14.445	107.4

Therefore we can also delete the regularity test for the Fisher information matrix in the selection algorithm in case of the CVS-model. In Figure 7 we show the minimal selection scores for $p = 3, \ldots, 20$, whereas in Table 10 we list the minimal selection scores $\alpha_{\min}(p)$, $p = 1, \ldots, 24$. Table 10 clearly shows the effect of the Fisher information matrix $\mathcal{F}_{25}$ extreme

Table 10: Minimal selection scores $\alpha_{\min}(p)$, $p = 1, \ldots, 24$, for the CVS-model.

p	1	2	3	4	5	6
$\alpha_{\min}(p)$	9.397×10^{-4}	2.042×10^{-2}	5.796×10^{-2}	1.096×10^{-1}	2.315×10^{-1}	3.299×10^{-1}
p	7	8	9	10	11	12
$\alpha_{\min}(p)$	4.340×10^{-1}	6.251×10^{-1}	8.018×10^{-1}	1.084	1.723	3.238
p	13	14	15	16	17	18
$\alpha_{\min}(p)$	6.336	17.13	37.31	59.37	2.224×10^{2}	4.301×10^{2}
p	19	20	21	22	23	24
$\alpha_{\min}(p)$	1.015×10^{3}	1.265×10^{5}	3.038×10^{5}	2.396×10^{31}	1.834×10^{34}	2.459×10^{108}

ill-conditioning. We see an extreme increase of the selection score from $3 \cdot 10^5$ at $p = 21$ to $2.4 \cdot 10^{31}$ at p=22. This is also reflected in Figure 6 which clearly shows that there is no reasonable approximation of $\log\big(\alpha_{\min}(p)\big)$, $p = 1, \ldots, 24$, by a regression line. However, a regression line makes sense for $p = 1, \ldots, 21$ as can be see from Figure 6. In Figure 7 we depict the minimal selection scores $\alpha_{\min}(p)$, $p = 2, \ldots, 19$, together with the approximating exponential functions

$$\alpha_{\min}(p) \approx 0.00670e^{0.580p}, \quad p = 1, \ldots, 19. \tag{32}$$

obtained from the regression line for $p = 1, \ldots, 19$ (solid grey line) and

$$\alpha_{\min}(p) \approx 0.00799e^{0.610p}, \quad p = 1, \ldots, 19. \tag{33}$$

obtained for $p = 2, 3, 4, 17, 18, 19$ (dashed grey line).

In Table 12 we present the sub-vector $\theta \in \mathbb{R}^{10}$ of q given by (27) with the minimal selection score together with the coefficients of variation. In Figure 8 we depict the classical sensitivities for the chosen sub-vector $\theta \in \mathbb{R}^{10}$ (black solid lines) and the sensitivities for the other 15 parameters (grey dashed lines). The right panel is a blow up for the sensitivities with values between -0.1 and 0.1. We see that the algorithm chooses predominantly parameters with larger sensitivities.

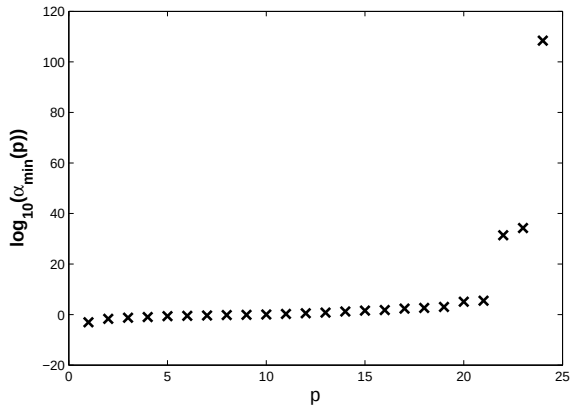


Figure 6: Logarithm of the minimal selection scores, $p = 1, \ldots, 24$, for the CVS-model.

Table 11: Time for computing $\theta \in \mathbb{R}^p$ with the minimal selection score on a laptop computer, once the 25×25 Fisher information matrix for the CVS-model has been computed.

p	1	2	3	4	5	6	7	8	9
time (sec)	0.0038	0.018	0.104	0.564	2.364	7.848	21.24	46.60	88.31
p	10	11	12	13	14	15	16	17	18
time (sec)	139.7	188.3	211.1	208.9	172.9	123.1	71.41	36.10	14.92
p	19	20	21	22	23	24			
time (sec)	5.137	1.382	0.329	0.093	0.0645	0.0005			

Table 12: Coefficients of variance for the sub-vector $\theta \in \mathbb{R}^{10}$ with the minimal selection score.

$\theta \in \mathbb{R}^{10}$	c_ℓ	α_ℓ	α_{r}	γ_ℓ	γ_{r}	ρ	q_{as}	V_{tot}	$R_{\mathrm{p}}^{\mathrm{rest}}$	$A_{\mathrm{pesk}}^{\mathrm{rest}}$
CV	0.465	0.087	0.172	0.353	0.622	0.114	0.454	0.290	0.248	0.213
$\alpha_{\min}(10)$	1.084									

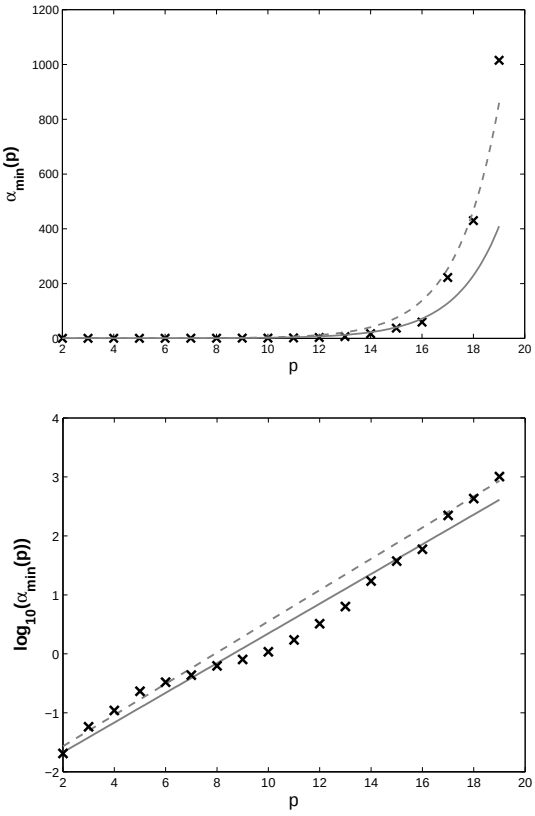


Figure 7: Minimal selection scores $\alpha_{\min}(p)$ (upper panel) and $\log_{10}(\alpha_{\min}(p))$ (lower panel), $p = 2, \ldots, 19$, for the CVS-model.

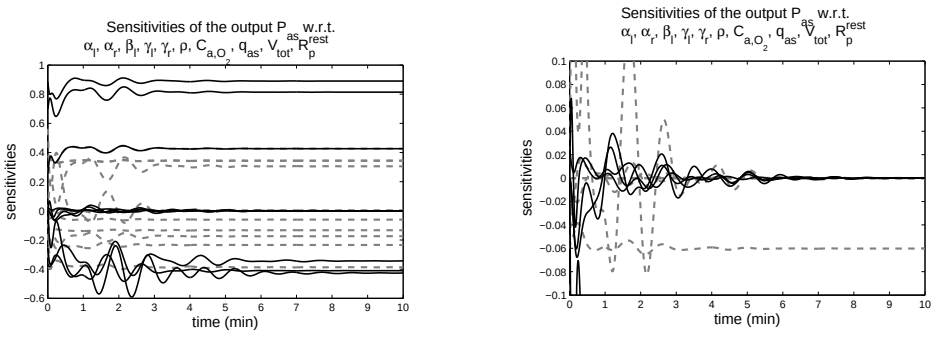


Figure 8: Classical sensitivities for the nominal values of $\theta \in \mathbb{R}^{10}$ with the minimal selection score (black solid lines) and classical sensitivities of the remaining 15 parameters (dashed grey lines).

6 Concluding Remarks

As we have noted, inverse problems for complex system models containing a large number of parameters are difficult. There is great need for quantitative methods to assist in posing inverse problems that will be well formulated in the sense of the ability to provide parameter estimates with quantifiable small uncertainty estimates. We have introduced and illustrated use of such an algorithm that requires prior local information about ranges of admissible parameter values and initial values of interest along with information on the error in the observation process to be used with the inverse problem. These are needed in order to implement the sensitivity/Fisher matrix based algorithm.

Because sensitivity of a model with respect to a parameter is fundamentally related to the ability to estimate the parameter, and because sensitivity is a local concept, we observe that the pursuit of a global algorithm to use in formulating parameter estimation or inverse problems is most likely a quest that will go unfulfilled.

Acknowledgements. This research was supported in part by Grant Number R01AI071915-07 from the National Institute of Allergy and Infectious Diseases and in part by the Air Force Office of Scientific Research under grant number FA9550-09-1-0226. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIAID, the NIH or the AFOSR.

References

[1] B.M. Adams, Non-parametric parameter estimation and clinical data fitting with a model of HIV infection, Ph.D. Thesis, North Carolina State University, Raleigh, 2005.

[2] B.M. Adams, H.T. Banks, M. Davidian, H. Kwon, H.T. Tran, S.N. Wynne, and E.S. Rosenberg, HIV dynamics: modeling, data analysis, and optimal treatment protocols, *J. Comp. and Appl. Math.* **184** (2005), 10 – 49.

[3] B.M. Adams, H.T. Banks, H.T. Tran, and H. Kwon, Dynamic multidrug therapies for HIV: optimal and STI control approaches, *Math. Biosci. and Engr.* **1** (2004), 223 – 241.

[4] B.M. Adams, H.T. Banks, M. Davidian, and E.S. Rosenberg, Model fitting and prediction with HIV treatment interruption data, CRSC-TR05-40, NCSU, October 2005; *Bull. Math. Biol.* **69** (2007), 563 – 584.

[5] D.T. Anh, M.P. Bonnet, G. Vachaud, C.V. Minh, N. Prieur, L.V. Duc, and L.L. Anh, Biochemical modeling of the Nhue River (Hanoi, Vietnam): practical identifiability analysis and parameter estimation, *Ecol. Model.* **193** (2006), 182 – 204.

[6] K.J. Astrom and P. Eykhoff, System identification – A survey, *Automatica* **7** (1971), 123 – 162.

[7] H.T. Banks, M. Davidian, J.R. Samuels, and K.L. Sutton, An inverse problem statistical methodology summary, Center for Research in Scientific Computation Technical Report CRSC-TR08-1, NCSU, January, 2008; in *Mathematical and Statistical Estimation Approaches in Epidemiology*, (G. Chowell et. al., eds.), Springer, New York, 2009, pp. 249 – 302.

[8] H.T. Banks, M. Davidian, S. Hu, G.M. Kepler, and E.S. Rosenberg, Modeling HIV immune response and validation with clinical data, CRSC-TR07-09, March, 2007; *J. Biological Dynamics* **2** (2008), 357 – 385.

[9] H.T. Banks, S. Dediu, and S.E. Ernstberger, Sensitivity functions and their uses in inverse problems, *CRSC Technical Report*, CRSC-TR07-12, NCSU, July, 2007; *J. Inverse and Ill-posed Problems* **15** (2007), 683 – 708.

[10] H.T. Banks, S. Dediu, S.L. Ernstberger, and F. Kappel, Generalized sensitivities and optimal experimental design, Center for Research in Scientific Computation Technical Report CRSC-TR08-12, NCSU, September, 2008, Revised, November, 2009; *J. Inverse and Ill-Posed Problems* **18** (2010), 25 – 83.

[11] H.T. Banks, S.L. Ernstberger, and S.L. Grove, Standard errors and confidence intervals in inverse problems: sensitivity and associated pitfalls, *J. Inverse and Ill-posed Problems* **15** (2007), 1 – 18.

[12] H.T. Banks and B.G. Fitzpatrick, Inverse problems for distributed systems: statistical tests and ANOVA, LCDS/CCS Rep. 88 – 16, July, 1988, Brown University; *Proc. International Symposium on Math. Approaches to Envir. and Ecol. Problems*, Springer Lecture Notes in Biomath. **81** (1989), 262 – 273.

[13] H.T. Banks and B.G. Fitzpatrick, Statistical methods for model comparison in parameter estimation problems for distributed systems, CAMS Tech. Rep. 89-4, September 1989, University of Southern California; *J. Math. Biol.* **28** (1990), 501 – 527.

[14] H.T. Banks, K. Holm, and D. Robbins, Standard error computations for uncertainty quantification in inverse problems: Asymptotic theory vs. bootstrapping, Center for Research in Scientific Computation Technical Report CRSC-TR09-13, NCSU, June 2009; Revised August 2009; *Arabian Journal for Science and Engineering: Mathematics (AJSE-Mathematics)*, submitted.

[15] H.T. Banks and J.R. Samuels, Jr., Detection of cardiac occlusions using viscoelastic wave propagation, CRSC-TR08-23, December 2008; *Advances in Applied Mathematics and Mechanics* **1** (2009), 1 – 28.

[16] J.J. Batzel, F. Kappel, D. Schneditz, and H.T. Tran, *Cardiovascular and Respiratory Systems: Modeling, Analysis and Control*, Frontiers in Applied Mathematics **FR34**, SIAM, Philadelphia, 2006.

[17] E.J. Bedrick and C.L. Tsai, Model selection for multivariate regression in small samples, *Biometrics* **50** (1994), 226 – 231.

[18] R. Bellman and K.M. Astrom, On structural identifiability, *Math. Biosci.* **7** (1970), 329 – 339.

[19] R. Bellman and R. Kalaba, *Quasilinearization and Nonlinear Boundry Value Problems*, American Elsevier, New York, 1965.

[20] S. Bonhoeffer, M. Rembiszewski, G.M. Ortiz, and D.F. Nixon, Risks and benefits of structured antiretroviral drug therapy interruptions in HIV-1 infection, *AIDS* **14** (2000), 2313 – 2322.

[21] H. Bozdogan, Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions, *Psychometrika* **52** (1987), 345 – 370.

[22] H. Bozdogan, Akaike's Information Criterion and recent developments in information complexity, *Journal of Mathematical Psychology* **44** (2000), 62 – 91.

[23] M. Burth, G.C. Verghese, and M. Vélez-Reyes, Subset selection for improved parameter estimation in on-line identification of a synchronous generator, *IEEE T. Power Syst.* **14** (1999), 218 – 225.

[24] K.P. Burnham and D.R. Anderson, *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, Springer, Berlin, Heidelberg, New York, 2002.

[25] K.P. Burnham and D.R. Anderson, Multimodel inference: Understanding AIC and BIC in model selection, *Sociological Methods and Research* **33** (2004), 261 – 304.

[26] D.S. Callaway and A.S. Perelson, HIV-1 infection and low steady state viral loads, *Bulletin of Mathematical Biology* **64** (2001) 29 – 64.

[27] A. Cintrón-Arias, C. Castillo-Chávez, L.M.A. Bettencourt, A.L. Lloyd, and H.T. Banks, The estimation of the effective reproductive number from disease outbreak data, Tech Rep CRSC-TR08-08, NCSU, April 2008; *Math. Biosci. Engr.* **6** (2009), 261 – 283.

[28] A. Cintrón-Arias, H.T. Banks, A. Capaldi, and A.L. Lloyd, A sensitivity matrix based methodology for inverse problem formulation, Tech Rep CRSC-TR09, NCSU, April 2009; *J. Inverse and Ill-posed Problems* **17** (2009), 545 – 564.

[29] C. Cobelli and J.J. DiStefano III, Parameter and structural identifiability concepts and ambiguities: a critical review and analysis, *Am. J. Physiol.* **239** (1980), R7 – R24.

[30] M. Davidian and D.M. Giltinan, *Nonlinear Models for Repeated Measurement Data*, Chapman & Hall, Boca Raton, 1995.

[31] B. Efron and R.J. Tibshirani, *An Introduction to the Bootstrap*, Chapman & Hall / CRC, Boca Raton, 1998.

[32] H.W. Engl, C. Flamm, P. Kügler, J. Lu, S. Müller, and P. Schuster, Inverse problems in system biology, *Inverse Problems* **25** (2009), 123014(51pp).

[33] P. Eykhoff, *System Identification: Parameter and State Estimation*, Wiley & Sons, New York, 1974.

[34] M. Eslami, *Theory of Sensitivity in Dynamic Systems: an Introduction*, Springer-Verlag, New York, NY, 1994.

[35] N.D. Evans, L.J. White, M.J. Chapman, K.R. Godfrey, and M.J. Chappell, The structural identifiability of the susceptible infected recovered model with seasonal forcing, *Math. Biosci.* **194** (2005), 175 – 197.

[36] M. Fink, myAD: fast automatic differentiation code in MATLAB, 2006; http://gosh.gmxhome.de/

[37] M. Fink, A. Attarian and H. Tran, Subset selection for parameter estimation in an HIV model, *Proc. Applied Math. and Mechanics* **7** (2008), 1121501 – 1121502.

[38] M. Fink, J.J. Batzel, and H. Tran, A respiratory system model: parameter estimation and sensitivity analysis, *Cardiovasc Eng.* **8** (2008), 120 – 134.

[39] K. Glover and J.C. Willems, Parametrizations of linear dynamical systems: Canonical forms and identifiability, *IEEE Trans. Automat. Contr.* **AC-19** (1974), 640 – 645.

[40] G.H. Golub and C.F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, 1996.

[41] C.M. Hurvich and C.L. Tsai, Regression and time series model selection in small samples, *Biometrika* **76** (1989), 297 – 307.

[42] A. Holmberg, On the practical identifiability of microbial growth models incorporating Michaelis-Menten type nonlinearities, *Math. Biosci.*, **62** (1982), 23 – 43.

[43] L.E. Jones and A.S. Perelson, Opportunistic infection as a cause of transient viremia in chronically infected HIV patients under treatment with HAART, *Bull. Math. Biol.*, **67** (2005), 1227 – 1251.

[44] R.E. Kalman, Mathematical description of linear dynamical systems, *SIAM J. Control*, **1** (1963), 152 – 192.

[45] F. Kappel and R.O. Peer, A mathematical model for fundamental regulation processes in the car- diovascular system, *J. Math. Biology* **31** (1993), 611 – 631.

[46] D.G. Luenberger, *Optimization by Vector Space Methods*, John Wiley & Sons, New York, NY, 1969.

[47] A.K. Mehra and D.G. Lainiotis, *System Identification*, Academic Press, New York, 1976.

[48] I.M. Navon, Practical and theoretical aspects of adjoint parameter estimation and identifiability in meteorology and oceanography, *Dyn. Atmospheres and Oceans*, **27** (1997), 55 – 79.

[49] P. Nelson, N. Smith, S. Cuipe, W. Zou, G.S. Omenn, and M. Pietropaolo, Modeling dynamic changes in type 1 diabetes progression: Quantifying β-cell variation after the appearance of islet-specific autoimmune responses, *Math. Biosci. Eng.,* **6** (2009), 753 – 778.

[50] M.A. Nowak and C.R.M. Bangham, Population dynamics of immune responses to persistent viruses, *Science*, **272** (1996), 74 – 79.

[51] J.T. Ottesen, M.S. Olufsen, and J.K. Larsen, *Applied Mathematical Models in Human Physiology*, Monographs on Mathematical Modeling and Computation, **MM09**, SIAM, Philadelphia, 2004.

[52] A.S. Perelson and P.W. Nelson, Mathematical analysis of HIV-1 dynamics in vivo, *SIAM Review*, **41** (1999), 3 – 44.

[53] J.G. Reid, Structural identifiability in linear time-invariant systems, *IEEE Trans. Automat. Control*, **22** (1977), 242 – 246.

[54] A.P. Sage and J.L. Melsa, *System Identification*, Academic Press, New York, 1971.

[55] G.A.F. Seber and C.J. Wild, *Nonlinear Regression*, John Wiley & Sons, Chichester, 2003.

[56] J. Shao and D. Tu, *The Jackknife and Bootstrap*, Springer-Verlag, New York, 1995.

[57] K. Thomaseth and C. Cobelli, Generalized sensitivity functions in physiological system identification, *Ann. Biomed. Eng.*, **27**(5) (1999), 607 – 616.

[58] D. Wodarz and M.A. Nowak, Specific therapy regimes could lead to long-term immunological control of HIV, *Proc. National Academy of Sciences*, **96** (1999), 14464 – 14469.

[59] L.J. White, N.D. Evans, T.J.G.M. Lam, Y.H. Schukken, G.F. Medley, K.R. Godfrey and M.J. Chappell, The structural identifiability and parameter estimation of a multispecies model for the transmission of mastitis in diary cows, *Math. Biosci.*, **174** (2001), 77 – 90.

[60] H. Wu, H. Zhu, H. Miao, and A.S. Perelson, Parameter identifiability and estimation of HIV/AIDS dynamics models, *Bull. Math. Biol.*, **70** (2008), 785 – 799.

[61] X. Xia and C.M. Moog, Identifiability of nonlinear systems with application to HIV/AIDS models, *IEEE Trans. Automat. Contr.*, **48** (2003), 330 – 336.

[62] H. Yue, M. Brown, F. He, J. Jia, and D.B. Kell, Sensitivity analysis and robust experimental design of a signal transduction pathway system, *Int. J. Chem. Kinet.*, **40** (2008), 730 – 741.