System identifiability (symbolic computation) and parameter estimation (numerical computation)

Lilianne Denis-Vidal [a], Ghislaine Joly-Blanchard [b] and Céline Noiret [b]

[a] *University of Sciences and Tech. Lille, 59655 Villeneuve d'Ascq, France*
E-mail: denvid@attglobal.net
[b] *University of Technology Compiègne, BP 20 529, 60205 Compiègne, France*
E-mail: Ghislaine.Joly-Blanchard@utc.fr; Céline.Noiret@utc.fr

A system identification based on physical laws often involves a parameter estimation. Before performing an estimation problem, it is necessary to investigate its identifiability. This investigation leads often to painful calculations. Generally, the numerical computation of the parameters does not use these calculus. In this contribution we propose least-squares methods to link identifiability approaches with numerical parameter estimation.

Keywords: nonlinear autonomous and uncontrolled systems, estimation of parameters, pharmacokinetic applications

AMS subject classification: 93C15, 93C95

1. Introduction

This paper considers uncontrolled nonlinear dynamical systems which are described in state-space form as:

$$\Gamma^\theta \quad \begin{cases} \dot{x}(t,\theta) = f(x(t,\theta),\theta), & x(0,\theta) = x_0(\theta), \\ y(t,\theta) = h(x(t,\theta),\theta). \end{cases} \tag{1}$$

In fact, what follows is valid also in the case of controlled systems. But this paper is devoted mainly to the analysis of uncontrolled pharmacokinetic models. In (1), $x(t,\theta) \in \mathbb{R}^n$ and $y(t,\theta) \in \mathbb{R}^m$ denote the state variables and the output, respectively. It is assumed that $\theta \in \mathcal{U}_p$, an open subset of $\mathbb{R}^p$. The functions $f(x,\theta)$ and $h(x,\theta)$ are real and analytic on M (a connected open subset of $\mathbb{R}^n$ such that $x(t,\theta) \in M$ for every $\theta \in \mathcal{U}_p$ and every $t \in [0,T]$). The initial conditions are assumed not to be an equilibrium state of the system.

Definition 1.1. The model Γ^θ is said to be identifiable (respectively locally identifiable) at $\theta \in \mathcal{U}_p$ if for any $\tilde{\theta} \in \mathcal{U}_p$ (respectively there exists an open neighborhood W of $\theta \in \mathcal{U}_p$ such that for any $\tilde{\theta} \in W$), $\tilde{\theta} \neq \theta$, the systems Γ^θ and $\Gamma^{\tilde{\theta}}$ yield different outputs.

In fact, requiring identifiability at every $\theta \in \mathcal{U}_p$ is straightforward but not very useful. In most models there exist atypical points in $\mathcal{U}_p$ where the model is unidentifiable. The previous definitions should be generically extended so that:

Definition 1.2. The model Γ^θ is said to be globally (locally) structurally identifiable if it is globally (locally) identifiable at all $\theta \in \mathcal{U}_p$ except at the points of a subset of measure zero in $\mathcal{U}_p$.

There exists a substantial literature concerning the system identifiability [7], on the one hand, and the parameter estimation [4], on the other hand. But few contributions combine both items [5]. Indeed, the first step of the parameter estimation ought to consist in testing the identifiability of parameters, which gives relations between parameters and outputs. Thus it should be interesting to use these relations in the parameter estimation.

Following this idea, we have selected some specific methods testing the identifiability. In [3], we have shown how identifiability can be analyzed by using directly the differential algebraic relations between outputs and parameters. In this contribution, we explain how these relations can be used from a numerical point of view. Thus we propose several criteria using such relations.

This paper is organized as follows. In section 2, the relations stated in the identifiability analysis and tested in the numerical estimation are given. In section 3, two different approaches of the numerical estimation are presented. The first one corresponds to global optimization procedures, the second one to local optimization procedures, which improves the first estimation. The comparison of both approaches is realized through the solution of a pharmacokinetic model.

2. Differential or integro-differential relations between output and parameters

Ljung and Glad [5] have given a conceptual description of the solution of the identifiability problem with the tools of differential algebra. We have shown [3] an other approach based on the differential algebraic relations between y and θ, mentioned in the following subsection.

Example 2.1. The following example allows the illustration of the various concepts:

$$
\begin{cases}
\dfrac{\mathrm{d}x_1}{\mathrm{d}t} = k_{12}(x_2 - x_1) - k_v \dfrac{x_1}{1 + x_1}, \\[2mm]
\dfrac{\mathrm{d}x_2}{\mathrm{d}t} = k_{21}(x_1 - x_2), \\[2mm]
y = x_1.
\end{cases}
\tag{2}
$$

The parameters are $\theta = (k_{12}, k_{21}, k_v)$ and the admissible set $\mathcal{U}_p = (\mathbb{R}_+^*)^3$.

2.1. Differential system of observation

In this section, the initial conditions are not taken into account. The elimination of state variables in (1) leads to a system of the form

$$R(y, \theta) = 0 \tag{3}$$

called the differential system of observation. The size of this system is the number of observations

$$R_j(y, \theta) = 0, \quad 1 \leqslant j \leqslant m, \tag{4}$$

where $R_j(y, \theta)$ is a differential polynomial in y and θ. More precisely,

$$R_j(y, \theta) = \sum_{1 \leqslant i \leqslant l+1} R_{i,j}(y)\rho_i(\theta), \quad j = 1, \dots, m, \tag{5}$$

where $R_{i,j}(y)$ is polynomial with respect to $(y, \dot{y}, \dots, y^{(r_j)})$. The integers r_j and l depend on state variable elimination and $\rho_i(\theta)$, $i = 1, \dots, l+1$, are algebraic polynomials in θ. Equation (5) can be rewritten as

$$\sum_{i=1}^{l} \phi_{i,j}(y)g_i(\theta) = \phi_{l+1,j}(y), \quad j = 1, \dots, m, \tag{6}$$

where $(g_i(\theta))_{1 \leqslant i \leqslant l}$ are rational in θ, $(\phi_{i,j})_{1 \leqslant i \leqslant l, 1 \leqslant j \leqslant m}$ are polynomial with respect to $(y, \dot{y}, \dots, y^{(r_j)})$ and $\phi_{l+1,j} \neq 0$ for an index $1 \leqslant j \leqslant m$.

The differential algebraic relations (6) between output and parameters lead to identifiability results. Moreover we have written a software which does an exhaustive analysis of the identifiability problem from the data of the functions f and h. This software uses an algorithm implemented within a package in MAPLE VIII by Boulier et al. [1] that returns all the solutions: a general case and particular cases.

In the case of example 2.1, the software returns two particular cases ($k_{12} = 0$ and $k_v = 0$) incompatible with the definition of $\mathcal{U}_p$ and the general case ($1 + y \neq 0$):

$$\ddot{y}(1 + y)^2 + (k_{12} + k_{21})\dot{y}(1 + y)^2 + k_{21}k_v y(1 + y) + k_v \dot{y} = 0, \tag{7}$$

that corresponds to equation (6). Using equation (7), the software checks that, if $y(\cdot, \theta) = y(\cdot, \tilde{\theta})$, then $g_i(\theta) = g_i(\tilde{\theta})$, $i = 1, 2, 3$, with:

$$g_1(\theta) = k_{12} + k_{21}, \qquad g_2(\theta) = k_{21}k_v, \qquad g_3(\theta) = k_v. \tag{8}$$

This gives $\theta = \tilde{\theta}$ and the structural global identifiability of the model.

2.2. Integro-differential system of observation

When initial conditions play a role in the identifiability procedure, we have presented a method [2] for taking them into account. But this leads only to identifiability results and not to a numerical estimation. Another way of considering these initial conditions is the integration from 0 to t of equation (6) or some equivalent relations which

could be integrated more easily. Sometimes the integration of the equations of the system (1) followed by the elimination of the state variables would be better. The aim of this computation is not only to facilitate the identifiability analysis by the introduction of the initial conditions but also to reduce the highest order of the time derivative for the numerical estimation as we will see later.

The integration will be done as far as possible and the so-obtained relations will be called the integro-differential system of observation:

$$T(y, \theta) = 0. \tag{9}$$

This system is an integro-differential system because there are some potential output derivatives left. Since equation (9) can be obtained by integrating equation (6) and since equation (6) is linear with respect to some parameter combinations, $T(y, \theta)$ is also linear with respect to some parameter combinations. The introduction of the initial conditions can give additional combinations, which is illustrated by the following example. We denote by $\{\gamma_1(\theta), \ldots, \gamma_{\hat{l}}(\theta)\}$ the parameter combinations appearing in $T(y, \theta)$.

In the case of example 2.1, we introduce the following initial conditions:

$$x_1(0) = x_{10}, \qquad x_2(0) = 0. \tag{10}$$

Either we integrate twice the following relation, equivalent to equation (7)

$$\ddot{y} = -(k_{12} + k_{21})\dot{y} - k_{21}k_v \frac{y}{1+y} - k_v \frac{\dot{y}}{(1+y)^2}, \tag{11}$$

or we integrate once the equations of (2), and then we eliminate x_2. In both cases, the following relation is obtained after some computation:

$$y(t) - x_{10} = -(k_{12} + k_{21}) \int_0^t y(u) \, du - k_v \int_0^t \frac{y(u)}{1+y(u)} \, du + k_{21}x_{10}t$$
$$- k_{21}k_v \int_0^t \int_0^u \frac{y(s)}{1+y(s)} \, ds \, du. \tag{12}$$

Equation (12) does not contain any output derivatives. There are four parameter combinations:

$$\gamma_1(\theta) = k_{12} + k_{21}, \qquad \gamma_2(\theta) = k_{21}k_v, \qquad \gamma_3(\theta) = k_v, \qquad \gamma_4(\theta) = k_{21}, \tag{13}$$

and this shows again the structural global identifiability of the model since the relations (13) contain combinations (8).

3. Numerical estimation of parameters

The *exact* observation, corresponding to the *exact* value $\bar{\theta}$, is more or less perturbed by noise. Thus, equations (6) and (9) are not satisfied by the noisy output. Moreover the observation is not carried out at any time and the given data are a set of measurements:

$$z_j := y(t_j; \bar{\theta}) + \varepsilon_j, \quad j = 1, 2, \ldots, q. \tag{14}$$
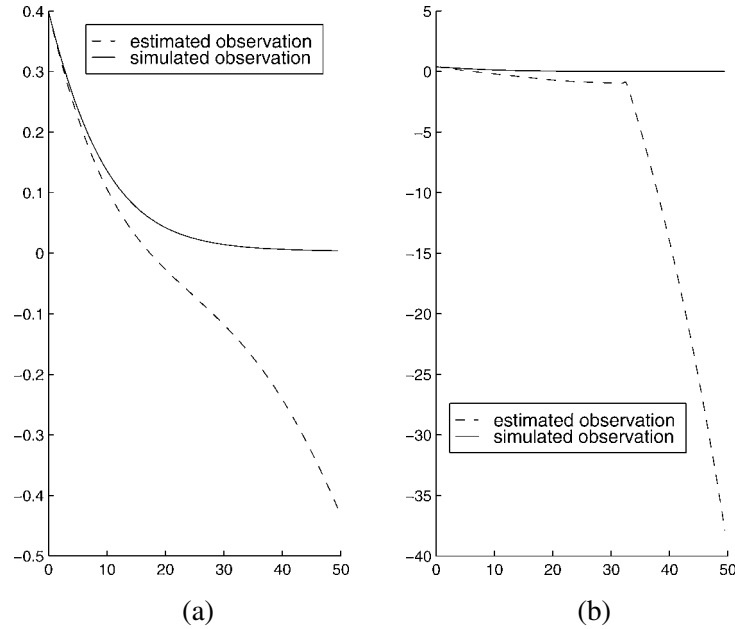
Figure 1. (a) $\theta_{io} = (0.018, 0.117, -0.08)$, (b) $\theta_{io} = (0.12, 0.07, -0.17)$.

In order to simplify the statement, the observation is assumed to be scalar ($m = 1$).

The various estimation procedures are applied to example 2.1. In order to compare their performances, the noisy data have been obtained by simulation ($\bar{\theta} = (0.01, 0.12, 0.02)$) in considering $\varepsilon_j \sim \mathcal{N}(0, 10^{-3})$, then $\varepsilon_j \sim \mathcal{N}(0, 10^{-2})$. In the following, numerical results will be presented in figures in which the solid line presents the simulated noisy observation and the dashed line presents the output corresponding to estimated parameter values. The left part of each figure is relative to the standard deviation 10^{-3} and the right part is relative to the standard deviation 10^{-2}. The scale and the range of the X-axis are identical for every figure. On the other hand, the scale and the range of the Y-axis are specific for some figures (figures 1 and 2(b)) because the results presented in these figures are out of a realistic range. But when the numerical results are satisfactory, the scale and the range of the Y-axis are identical in the concerned figures.

3.1. Global optimization

The aim of this section is the estimation of the parameters directly from equation (6) or equation (9). Both relations are assumed to prove model identifiability; otherwise the numerical parameter estimation elaborated from equation (6) or equation (9) would not be valid.

Output derivatives are involved in equation (6), sometimes even in equation (9), and they have to be estimated from a noisy observation. Different methods are available to estimate these derivatives in (t_j): local approximation polynomials, cubic splines, etc. However these approximations are sensitive to noise, especially if the derivative
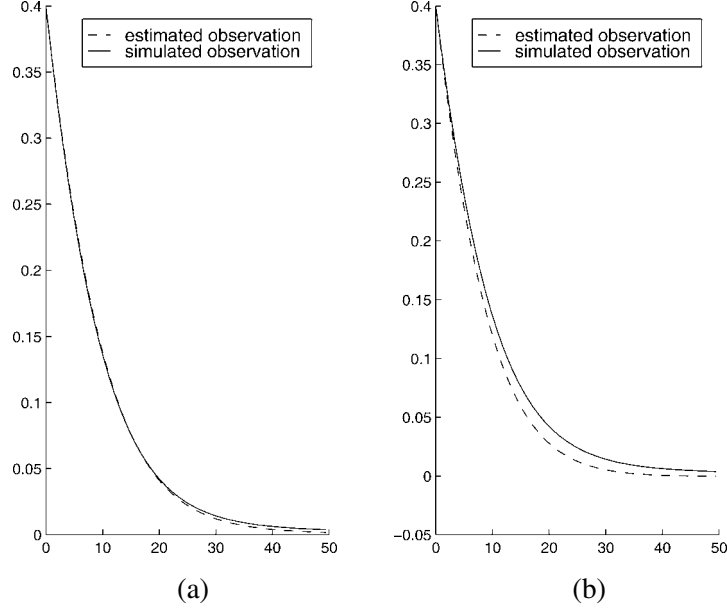
Figure 2. (a) $\theta_{\text{int}} = (0.004, 0.12, 0.025)$, (b) $\theta_{\text{int}} = (-0.003, 0.15, -0.06)$.

order is large. Moreover, they are not accurate near the bounds of the time interval of observation.

We consider derivatives to be estimated in $\{t_{n_0}, \ldots, t_{n_1}\}$ $(n_1 - n_0 \ggg p)$ and we denote by:

- $\{P_{\text{out}}(z_j, \theta)\}_{n_0 \leqslant j \leqslant n_1}$ the values of the polynomial corresponding to the approximation of $\sum_{i=1}^{l} \phi_i(y) g_i(\theta) = \phi_{l+1}(y)$ (equation (6)) in $(t_j)_{n_0 \leqslant j \leqslant n_1}$,
- $\{P_{\text{int}}(z_j, \theta)\}_{n_0 \leqslant j \leqslant n_1}$ the values of the polynomial corresponding to the approximation of $T(y, \theta)$ (equation (9)) in which integrals have also been estimated.

The integers n_0 and n_1 could possibly be different in both approximations especially if the integro-differential polynomial of observation does not contain any derivatives.

Since outputs are noisy and derivatives estimated, values of $\{g_1(\theta), \ldots, g_l(\theta)\}$ cannot be exact. Therefore the following minimization problem:

$$\min_{(g_1(\theta), \ldots, g_l(\theta))} \sum_{j=n_0}^{j=n_1} \left(\mathcal{P}_{\text{out}}(z_j, \theta) \right)^2 \tag{15}$$

has to be solved in order to estimate $\{g_1(\theta), \ldots, g_l(\theta)\}$.

In the same way, the following minimization problem:

$$\min_{(\gamma_1(\theta), \ldots, \gamma_{\hat{l}}(\theta))} \sum_{j=n_0}^{j=n_1} \left(\mathcal{P}_{\text{int}}(z_j, \theta) \right)^2 \tag{16}$$

has to be solved in order to estimate $\{\gamma_1(\theta), \ldots, \gamma_{\hat{l}}(\theta)\}$.

In both cases polynomials are linear with respect to involved parameter combinations, so a QR algorithm is applied to solve minimization problems (solutions of (15) and (16) are assumed to be unique). Moreover, the identifiability of the model is assumed to be shown from equation (6) (respectively equation (9)). Thus, unique parameter values θ are drawn from $\{g_1(\theta), \ldots, g_l(\theta)\}$ (respectively $\{\gamma_1(\theta), \ldots, \gamma_{\hat{l}}(\theta)\}$). Let us remark that l (respectively $\hat{l}$) is greater or equal to p ($\theta \in \mathbb{R}^p$ and that $\{g_1(\theta), \ldots, g_l(\theta)\}$ (respectively $\{\gamma_1(\theta), \ldots, \gamma_{\hat{l}}(\theta)\}$) are not exact. Consequently, another least-squares problem has possibly to be solved for estimating θ.

Application to example 2.1

The first and the second time derivatives of y have to be estimated to get $\mathcal{P}_{\text{out}}(z_j, \theta)$. The solution of (15) is presented in figure 1. The bad results concerning both estimated parameters (θ_{io}) and the estimated output are due mainly to the approximation of derivatives of a noisy output.

The achieving of $\mathcal{P}_{\text{int}}(z_j, \theta)$ requires only approximate values of integrals of y from 0 to t, obtained by a trapezoidal rule. The results are given in figure 2. This seems more satisfactory relatively to the estimated output.

The parameter vector (θ_{io}) (respectively (θ_{int})) obtained from solving (15) (respectively (16)) may loose accuracy due to three factors:

- the computation of $\mathcal{P}_{\text{out}}$ (respectively $\mathcal{P}_{\text{int}}$),
- the solution of the least-squares problem,
- the calculation of (θ_{io}) from $\{g_1(\theta), \ldots, g_l(\theta)\}$ (respectively (θ_{int}) from $\{\gamma_1(\theta), \ldots, \gamma_{\hat{l}}(\theta)\}$.

The main difference between both solutions comes from the first error source since only the computation of $\mathcal{P}_{\text{out}}$ requires derivative estimations. The condition number of both least-squares problems are comparable. Lastly, there is more information in γ_i ($1 \leqslant i \leqslant 4$) than in g_i ($1 \leqslant i \leqslant 3$). These remarks agree with results being better by solving the minimization problem (16). However, the relative errors are not very small, which has led to implement local optimization methods to improve the accuracy.

3.2. Local optimization

Since the estimated parameters obtained in the previous section are not accurate enough, they can be improved by a nonlinear programming algorithm. Indeed, if there is no information available about magnitude order of parameter values, they can be used as the starting point of a local optimization procedure.

Penalized criterion. The estimated parameters obtained in section 3.1 may be outside the admissible set $\mathcal{U}_p$. Consequently, the criterion (16) can be penalized by the contraints corresponding to $\mathcal{U}_p$. The obtained criterion is no longer linear with respect to the parameter combinations $\{\gamma_1(\theta), \ldots, \gamma_{\hat{l}}(\theta)\}$.
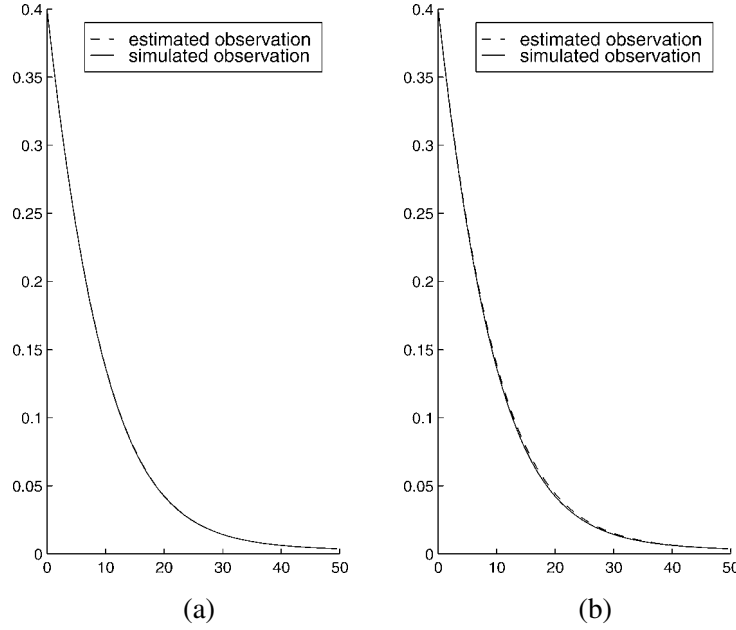
Figure 3. (a) $\theta_{\text{pid}} = (0.113, 0.118, 0.015)$, (b) $\theta_{\text{pid}} = (0.008, 0.12, 0.027)$.

In the case of example 2.1, the definition of $\mathcal{U}_p = (\mathbb{R}_+^*)^3$, leads to a penalization by $\|\theta^-\|^2$, where $\theta_i^- = \min(0, \theta_i)$ ($1 \leqslant i \leqslant 3$):

$$\min_{(\gamma_1(\theta),\ldots,\gamma_{\tilde{l}}(\theta))} \left(\sum_{j=n_0}^{j=n_1} \left(\mathcal{P}_{\text{int}}(z_j, \theta) \right)^2 + \varepsilon \|\theta^-\|^2 \right). \tag{17}$$

This criterion has been minimized by Levenberg–Marquardt algorithm, with the starting point given by figure 2. The results are given in figure 3. They are very satisfactory relatively to parameter values (θ_{pid}) and the reconstructed output. Once more, the relations given by identifiability analysis are involved in the estimation procedure.

Weighting method: linear matrix inequality (LMI) approach. In this part, we are trying to improve the previous estimated parameters θ_{pid} by using a classical least-squares objective function based on residuals at times t_i. But we modify the least-squares objective function by means of a weighting operator in order to improve the condition number of the estimation problem:

$$\min_{\theta} J_W(\theta), \quad \text{where } J_W(\theta) = \sum_{i=n_0}^{n_1} w_i \left(y(t_i, \theta) - z_i \right)^2. \tag{18}$$

Ouarit proposes in [6] to formulate this problem in term of LMI. The problem consists in looking for the weight matrix $W = \text{diag}(w_{n_0}, \ldots, w_{n_1})$ by considering the following
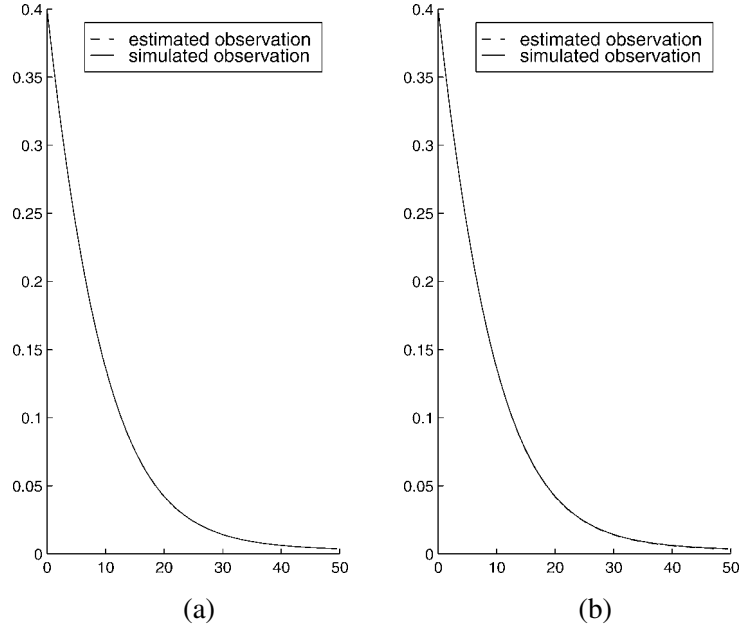
Figure 4. (a) $\theta_{\text{lmi}} = (0.011, 0.119, 0.017)$, (b) $\theta_{\text{lmi}} = (0.008, 0.123, 0.026)$.

criterion:

$$
\min_{w_{n_0}, \ldots, w_{n_1}, t, r} \quad t + \varepsilon r
$$

$$
\text{subject to} \quad
\begin{bmatrix}
tI & G_W - I \\
(G_W - I)^{\text{T}} & tI
\end{bmatrix} > 0,
$$

$$
\begin{bmatrix}
rI & w \\
w^{\text{T}} & 1
\end{bmatrix} > 0,
$$

$$
W > 0,
$$

(19)

where G_W is the Gauss–Newton approximation of the Hessian matrix and the vector w is defined by $w = (w_{n_0}, \ldots, w_{n_1})^{\text{T}}$. This leads to the following algorithm:

Initialization $\theta_0 = \theta_{\text{pid}}$.

Step k 1. Calculation of the weighting matrix W by solving (19) for $\varepsilon = 10^{-3}$.

2. Minimization of the criterion (18).

In the case of example 2.1, problem (19) comes within the linear programming and is solved by using the function minmax of the toolbox LMI of Matlab 4.2. Criterion (18) is minimized by Levenberg–Marquard algorithm. The results presented in figure 4 show a slight improvement of the estimated values of the parameters, but this algorithm is time consuming.

4. Conclusion

The theoretical identifiability analysis, performed with the help of symbolic computation software, leads to relations between inputs, outputs and parameters. This contribution starts from these relations to elaborate some criteria whose minimization gives numerical estimations of parameters even with noisy outputs. Thanks to this new procedure, we obtain quadratic criteria with respect to some parameter combinations. Therefore, this supplies parameter estimations which do not necessitate a priori information about these parameters. One of the criteria (integro-differential criterion) gives a first approximation that is significantly improved by a penalization term corresponding to the constraints of the admissible set. Finally, the last results have been slightly improved by a classical criterion of parameter estimation. This algorithm optimizes the condition number of the least-squares problem but it is time consuming. Other practical examples have allowed to test this procedure. It is especially interesting in the case where identifiability analysis leads to relations between inputs, outputs and parameters including derivatives of weak order, or even no derivative.

References

[1] F. Boulier, D. Lazard, F. Ollivier and M. Petitot, Representation for the radical of a finitely generated differential ideal, in: *Proc. of ISSAC'95, Internat. Symposium on Symbolic and Algebraic Computation*, Montréal, Canada, 1995, pp. 158–166.

[2] L. Denis-Vidal, G. Joly-Blanchard and C. Noiret, Some effective approaches to check the identifiability of uncontrolled nonlinear systems, Math. Comput. Simulation 57 (2001) 35–44.

[3] L. Denis-Vidal, G. Joly-Blanchard, C. Noiret and M. Petitot, An algorithm to test identifiability of non-linear systems, in: *Proc. of the 5th IFAC NOLCOS*, St Petersburg, Russia, 2001, pp. 174–178.

[4] L. Ljung, *System Identification: Practice for the User* (Prentice-Hall, Englewood Cliffs, NJ, 1989).

[5] L. Ljung and T. Glad, On global identifiability for arbitrary model parametrizations, Automatica 30(2) (1994) 265–276.

[6] M. Ouarit, J. P. Yvon and J. Henry, Optimal weighting design for distributed parameter system estimation, Optimal Control Appl. Methods 22 (2001) 37–49.

[7] E. Walter, Identifiability of state space models, in: *Lecture Notes Biomath.*, Vol. 46 (1982).