Computational aspects of a class of identification problems

Åström, Karl Johan

1969

Document Version:
Publisher's PDF, also known as Version of record

COMPUTATIONAL ASPECTS OF A CLASS OF IDENTIFICATION PROBLEMS

K. J. Åström

Department of Electrical Engineering
University of Southern California
Los Angeles, California 90007

and

Division of Automatic Control
Lund Institute of Technology
Lund 7, Sweden

Technical Report No. 69-14

October 1969

Supported by the National Institutes of Health
under Grant No. GM 16197-01.

COMPUTATIONAL ASPECTS OF A CLASS OF IDENTIFICATION PROBLEMS

1. Introduction

In practically all computational problems, there is a possibility to trade computing time against storage and also to reduce both computing time and storage by exploiting particular structural properties of the problem. In this note, we will look into these questions for a class of nonlinear identification problems. We will consider the identification of the parameter α in a system described by the equations

$$\frac{dx}{dt} = f(x, u, t, \alpha), \tag{1.1}$$

$$y = g(x, u, t, \alpha), \tag{1.2}$$

where u is the control variable and y the output. The identification problem is formulated as an optimization problem in Sec. 2 by introducing the loss function

$$v(\alpha) = \frac{1}{2}\int_0^t [y(s) - y_m(s)]^T dF(s)[y(s) - y_m(s)], \tag{1.3}$$

where y_m is the measured output signal. Different algorithms for minimizing the loss function are discussed in Sec. 3. Section 4 deals with the trade-off between computing time and storage when evaluating the gradient of the loss function.

The particular case of linear systems with constant coefficients and observations with equal spacing is discussed in Sec. 5. Section 6 is devoted to linear systems with a particular (companion) structure which leads to a significant reduction in computing time.

2. Formulation of the Problem

Consider a system described by the equations

$$\frac{dx}{dt} = f(x, u, t, \alpha), \tag{2.1}$$

$$y = g(x, u, t, \alpha), \tag{2.2}$$

where x is an n-dimensional state vector, u an r-dimensional vector of control variables, y a p-dimensional vector of output variables. The functions f and g depend on a set of parameters $\alpha = \mathrm{col}[\alpha_1, \alpha_2, \ldots, \alpha_m]$ whose values are unknown. The initial state of (2.1) is

$$x(0) = c. \tag{2.3}$$

It is assumed that c is known, and that the solution of (2.1) with initial value (2.3) exists over $(0, t)$. It is desired to estimate the unknown parameter α on the basis of measurements made during experiments of the system.

In an experiment an arbitrary control signal is chosen. The output of the system is measured continuously or at discrete times. The measured outputs are denoted by $y_m(t)$.

The parameter α should be determined in such a way that the following criterion is minimized,

$$v(\alpha) = \frac{1}{2} \int_0^t [y(s) - y_m(s)]^T dF(s)[y(s) - y_m(s)], \tag{2.4}$$

where F is a symmetric matrix. The difference $F(t) - F(s)$, $t > s$, determines the weight given to the measurements over the interval (t, s). We illustrate the physical interpretation of the function F by two examples.

Example 1. Consider the case of two outputs. If both outputs are measured continuously in time and the measurement of y is twice as accurate as y_2, we have

$$F(t) = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}. \tag{2.5}$$

Example 2. If measurements are available at discrete times $t = 0, 1, 2, \ldots$, only, and all measurements are equally important, we have

$$F(t) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \text{integer part of } t. \tag{2.6}$$

The criterion (2.4) is chosen rather arbitrarily. It can, however, be given a nice stochastic interpretation in the case when there are no disturbances acting on the system and the measurement errors are independent with a normal distribution.

3. Outline of Solution

The solution of the optimization problem will now be discussed. As the problem is nonlinear, it is necessary to use numerical methods. There are many algorithms available to find the maximum of a function of several variables. A large class of algorithms can be described by the recursive equation

$$\alpha^{n+1} = F[\alpha^n, \alpha^{n-1}, \ldots, V(\alpha^n), V(\alpha^{n-1}), \ldots, V_\alpha(\alpha^n), \ldots, V_{\alpha\alpha}(\alpha^n), \ldots], \tag{3.1}$$

where $\alpha^1, \alpha^2, \ldots,$ denotes the successive iterates, V_α is the gradient of V with respect to α, $V_{\alpha\alpha}$ denotes the matrix of second-order partial derivatives, etc.

Some methods require only the evaluation of the function V itself, other methods require evaluation of the gradient V_α, and sophisticated methods like the Newton-Raphson method [5],

$$\alpha^{n+1} = \alpha^n + V_{\alpha\alpha}^{-1}(\alpha^n) V_\alpha(\alpha^n), \tag{3.2}$$

require evaluation of the matrix of second-order derivation. In return for this, the Newton-Raphson method gives quadratic convergence near the extremum. There are also other methods, e.g., the Fletcher-Powell method, which give quadratic convergence without requiring evaluation of second-order derivatives.

Due to the particular structure of the loss function (2.4), it seems attractive to use a method which involves gradients to solve the identification problem. In the next section, we will discuss different methods to evaluate the gradients.

4. Evaluation of Derivatives of the Loss Function

In this section, we will discuss the evaluation of the gradient of the loss function. The same approach can also be used to evaluate derivatives of higher order. A straightforward method to evaluate the derivatives of the function V

defined by (2.4) has been described in [2] and [3]. The procedure is as follows. Differentiation of (2.4) with respect to α gives

$$V_\alpha = \int_0^t [y(s) - y_m(s)]^T dF(s) y_\alpha(s). \tag{4.1}$$

To obtain y_α, differentiate (2.2). Hence,

$$y_\alpha = g_\alpha + g_x x_\alpha. \tag{4.2}$$

Differentiation of (2.1) finally gives

$$\frac{d}{dt} x_\alpha = f_x x_\alpha + f_\alpha. \tag{4.3}$$

Combination of (4.1), (4.2), and (4.3) now gives a procedure for evaluating the gradient V_α. The main burden in this computation is the integration of the differential equation (4.3). Each parameter α_i requires the integration of a system of n first-order differential equations. The evaluation of x_α thus requires the solution of m n-dimensional differential equations.

A Different Method. We will now show that there is an alternative method of evaluating the gradient V_α which only requires the integration of one n-dimensional differential equation. This method will, however, require more storage. To obtain the alternative method, we will consider the differential equation (2.1) as a constraint which we take care of by the standard technique of Lagrangian multipliers. Let λ be an n-vector of continuously differentiable functions. As x satisfies the equation (2.1), the function V can be written as

$$V(\alpha) = \frac{1}{2} \int_0^t [y(s) - y_m(s)]^T dF(s) [y(s) - y_m(s)] - \int_0^t \lambda^T(s) \left[\frac{dx}{ds} - f(x, u, s, \alpha) \right] ds. \tag{4.4}$$

Partial integration gives

$$V(\alpha) = \frac{1}{2}\int_0^t [y(s) - y_m(s)]^T dF(s)[y(s) - y_m(s)] \tag{4.5}$$
$$- \lambda^T(t)x(t) + \lambda^T(0)x(0) + \int_0^t \left[\frac{d\lambda^T}{ds}\, x + \lambda^T f\right] ds.$$

Differentiation with respect to α gives

$$V_\alpha = \int_0^t [y(s) - y_m(s)]^T dF(s) g_\alpha(s) + \int_0^t [y(s) - y_m(s)]^T dF(s) g_x x_\alpha \tag{4.6}$$
$$- \lambda^T(t)x_\alpha(t) + \int_0^t \left[\frac{d\lambda^T}{ds}\, x_\alpha + \lambda^T f_x x_\alpha\right] ds.$$

Up to now λ has been arbitrary. Now choose λ such that

$$\lambda(t) = 0, \tag{4.7}$$

$$\frac{d\lambda}{dt} + f_x^T \lambda + g_x^T \frac{dF}{dt}[y - y_m] = 0 \quad \text{when } \frac{dF}{dt} \text{ exists},$$

$$\lambda(t+) - \lambda(t-) + g_x^T [F(t+) - F(t-)][y(t) - y_m(t)] \quad \text{at discontinuities of } F.$$

It then follows from (4.6) that the gradient V_α can be written as

$$V_\alpha = \int_0^t [y(s) - y_m(s)]^T dF(s) g_\alpha + \int_0^t \lambda^T f_\alpha ds. \tag{4.8}$$

We thus get the following procedure for calculating the gradient:

(1) Integrate (2.1) in the forward direction with initial condition.

(2) Integrate the differential equation (4.7) for λ backwards starting with $\lambda(t) = 0$ and taking into account the jumps which occur at the jumps of F.

(3) Evaluate the gradient by quadrature using (4.8).

With this approach, it is thus sufficient to integrate one n-dimensional differential equation in order to obtain the gradient. Notice, however, that it is required to store both $\{x(s),\ 0 \le s \le t\}$ and $\{\lambda(s),\ 0 \le s \le t\}$.

5. Linear Systems with Observations at Equally Spaced Instants of Time

We will now turn to linear systems with constant coefficients, i.e., systems described by

$$\frac{dx}{dt} = Ax + By, \tag{5.1}$$

$$y = cx, \tag{5.2}$$

where A, B, and C are matrices with constant elements. The matrix A is n x n, B is n x r, and C is p x n. It is also assumed that the output is observed at discrete instants of time. The unit of time is chosen so that observations are made at $t = 0, 1, 2, \ldots$. It is also assumed that the control signal is kept constant over the intervals $(n, n + 1]$. The control signal might change discontinuously at the observation intervals. The identification problem now has a particular structure which can be exploited in order to reduce the computations required. See [4]. We will illustrate the possible savings by considering the case when α is a scalar only. The equation for the derivative of x with respect to α now becomes

$$\frac{dx_\alpha}{dt} = Ax_\alpha + A_\alpha x + B_\alpha u. \tag{5.3}$$

In order to evaluate x_α, we thus have to integrate the equations

$$\frac{d}{dt}\begin{bmatrix} x \\ x_\alpha \end{bmatrix} = \begin{bmatrix} A & 0 \\ A_\alpha & A \end{bmatrix}\begin{bmatrix} x \\ x_\alpha \end{bmatrix} + \begin{bmatrix} B \\ B_\alpha \end{bmatrix} u. \tag{5.4}$$

As this equation is linear with constant coefficients, and as the control signal is kept constant over the intervals $(n, n + 1]$, we get

$$z(t + 1) = \Phi z(t) + \Gamma u(t), \tag{5.5}$$

where

$$z = \begin{bmatrix} x \\ x_\alpha \end{bmatrix}, \tag{5.6}$$

$$\Phi = \exp\begin{bmatrix} A & 0 \\ A_\alpha & A \end{bmatrix}. \tag{5.7}$$

The matrix Γ will depend on the way in which the control signal u is changed.

In drug kinetics experiments, the injection can often be considered as instantaneous. In such a case the control signal is zero except at the injection instant and the state variable will make a jump equal to $\begin{bmatrix} B \\ B_\alpha \end{bmatrix}$ times the injection amount. The situation can be modelled by (5.5) if we choose

$$\Gamma = \Phi\begin{bmatrix} B \\ B_\alpha \end{bmatrix}, \tag{5.8}$$

and let $u(t)$ denote the amount of drug which is injected at time t.

In problems where the control signal is kept constant over the sampling interval and is allowed to a finite jump at the sampling points, the model (5.5) is still valid if we choose

$$\Gamma = \left\{\int_0^t \exp\begin{bmatrix} A & 0 \\ A_\alpha & A \end{bmatrix} ds\right\}\begin{bmatrix} B \\ B_\alpha \end{bmatrix}, \tag{5.9}$$

and let $u(t)$ denote the actual value of the control signal at time t.

By exploiting equation (5.5) the amount of computation required to evaluate the gradient can be reduced considerably. The matrices Φ and Γ can be computed once and for all by using the series expansion

$$e^A = I + A + \frac{1}{2!}A^2 + \cdots + \frac{1}{n!}A^n + \cdots, \tag{5.10}$$

or by integrating the equation (5.4) over the interval $(0, 1)$. The values of x and the gradient x_α are then obtained simply by iterating (5.5).

As an alternative we could also dispose of the differential equation entirely and identify the coefficients of the difference equation (5.5). This is perfectly reasonable to do if the model will be used to design control strategies where the control actions will be taken only at $t = 0, 1, 2, \ldots$, and if we are

satisfied by knowing the state of the system at the sampling intervals only. If we are satisfied with a sampled model we can also use the discrete version at the computational scheme discussed in Sec. 4 in order to reduce the number of computations further. This would appear particularly attractive in the case of systems with long observation intervals.

6. Linear Systems of Companion Form

We will now specialize even further to linear systems with a particular structure. It is assumed that the matrix A is a companion form. The system can then be represented by the equation

$$\frac{dx}{dt} = \begin{bmatrix} -a_1 & 1 & 0 & \cdots & 0 \\ -a_2 & 0 & 1 & \cdots & 0 \\ \vdots & & & & \\ -a_{n-1} & 0 & 0 & \cdots & 1 \\ -a_n & 0 & 0 & \cdots & 0 \end{bmatrix} x + \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_{n-1} \\ b_n \end{bmatrix}, \tag{6.1}$$

$$y = [1 \quad 0 \quad 0 \quad \cdots \quad 0]x. \tag{6.2}$$

For simplicity, we consider the case of one input and one output only. Taking Laplace transforms we find that the input-output relation of the system given by (6.1) and (6.2) can be described by

$$Y(s) = \frac{b_1 s^{n-1} + b_2 s^{n-2} + \cdots + b_n}{s^n + a_1 s^{n-1} + \cdots + a_n} U(s) + \frac{x_1(0)s^{n-1} + x_2(0)s^{n-2} + \cdots + x_n(0)}{s^n + a_1 s^{n-1} + \cdots + a_n} \tag{6.3}$$

where Laplace transforms are denoted by capitals, i.e.,

$$Y(s) = \int_0^\infty e^{-st} y(t)dt. \tag{6.4}$$

To evaluate the gradient of the loss function (2.4), we need the derivatives $\partial Y/\partial a_i$ and $\partial Y/\partial b_i$. This can be done as follows. Differentiation of (6.3) gives

$$(6.5)\qquad \frac{\partial Y}{\partial a_i} = -\frac{s^{n-i}}{s^n + a_1 s^{n-1} + \cdots + a_n}\, Y(s).$$

This input-output relation can be represented by the state equation

$$(6.6)\qquad \frac{dz}{dt} = \begin{bmatrix} -a_1 & -a_2 & \cdots & -a_{n-1} & -a_n \\ 1 & 0 & & 0 & 0 \\ 0 & 1 & & 0 & 0 \\ \vdots & & & & \\ 0 & 0 & & 1 & 0 \end{bmatrix} z + \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} y,$$

$$\frac{\partial y}{\partial a_i} = z_i,$$

$$z(0) = 0.$$

Similarly, we find

$$(6.7)\qquad \frac{\partial Y}{\partial b_i} = \frac{s^{n-i}}{s^n + a_1 s^{n-1} + \cdots + a_n}\, U.$$

This input-output relation can be represented by the state equation

$$(6.8)\qquad \frac{dz}{dt} = \begin{bmatrix} -a_1 & -a_2 & \cdots & -a_{n-1} & -a_n \\ 1 & 0 & & 0 & 0 \\ 0 & 1 & & 0 & 0 \\ \vdots & & & & \\ 0 & 0 & & 1 & 0 \end{bmatrix} z + \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} u,$$

$$\frac{\partial y}{\partial b_i} = z_i,$$

$$z(0) = 0.$$

Hence, by exploiting the particular structure of the system we find that all derivatives with respect to the parameters b_i can be obtained from one single

differential equation. A comparison with Sec. 4 shows that in the general case it is necessary to integrate n equations.

If the observations are obtained at discrete times only, it is also possible to integrate over the interval (0, 1) only and then use the recursive equation developed in Sec. 5.

The ideas given in this section can also be exploited to calculate derivatives of higher order. See [1].

REFERENCES

1. Astrom, K. J., and T. Bahlin, "Numerical Identification of Linear Dynamic Systems from Normal Operating Records," Proc. IFAC Symposium on Self-Adaptive Control Systems, Teddington, September 1965.

2. Bellman, R., and R. Kalaba, Quasilinearization and Nonlinear Boundary Value Problems, American Elsevier Publishing Company, Inc., New York, 1965.

3. Buell, J., and R. Kalaba, "Quasilinearization and the Fitting of Nonlinear Models of Drug Metabolism to Experimental Kinetic Data," Math. Biosci., Vol. 5, 1969, pp. 121-132.

4. Buell, J., R. Kalaba, and E. Ruspini, Identification of Linear Systems Using Long Periods of Observation, University of Southern California, USCEE-363, June 1969.

5. Saaty, T. L., and J. Bram, Nonlinear Mathematics, McGraw-Hill Book Company, New York, 1969.