Identification of Continuous Systems – Practical Issues of Insensitivity to Perturbations

Janusz Kozłowski* and Zdzisław Kowalczuk

Gdańsk University of Technology, Faculty of Electronics, Telecommunications and Informatics, Department of Robotics and Decision Systems, Gdańsk, Poland
jk23@eti.pg.gda.pl, kova@eti.pg.gda.pl

Abstract. In this paper the issue of continuous systems estimation, insensitive to certain perturbations, is presented and discussed. Such an approach has rational advantages, especially when robust schemes are used to assist a target system responsible for industrial diagnostics. This requires that estimated model parameters are generated on-line, and their values are reliable and to a great extent accurate. Practical hints are suggested to challenge the consistency problem of estimates. Namely, the technique of instrumental variables can improve the asymptotic behavior of estimators. With a weighting mechanism, in turn, tracking the time-varying parameters of non-stationary processes is realistic. Yet, evident insensitivity to destructive outliers in the measurement data is guaranteed by the applied estimation routine in the sense of the least sum of absolute errors. Finally, premises for a proper selection of persistently exciting input signals, as well as the directions of further research are summarized in the paper.

Keywords: Continuous Models, Robust Identification, Least-Squares, Instrumental Variable, Least Absolute Errors.

1 Introduction

In practical automation systems, the idea of control and diagnostics of industrial processes is based on dedicated discrete-time or continuous-time models. Identification of such models, and examination of the estimated parameters, deliver valuable diagnostic information about the evolution of the respective supervised processes. Most often, detection of a hazardous situation in a monitored system, recognized from abrupt changes in the identified parameters, is not problematic. In the case of gradual changes in the model parameters, however, an early warning can protect the system from a potential serious damage.

It is of fundamental importance that the employed identification procedure generates reliable and accurate estimates of the employed model. Otherwise, the diagnostic system will either produce false alarms, or ignore possible hazardous outcomes. Useful identification methods insensitive to different perturbations (high frequency correlated noises, DC offsets in additive disturbances or destructive outliers) are discussed in this paper.

It is still open to question, whether discrete-time or continuous-time models should be involved in diagnostic procedures utilizing algorithms for identification or change detection [3]. On the one hand, discrete-time models (with the regression data usually represented by shifted samples of registered input-output data) can easily be handled numerically, and parameter estimation of such models is straightforward. On the other hand, such 'mathematical' parameters are dimensionless, have no physical interpretation, and depend on the applied sampling frequency. When using the continuous-time models, in turn, the intuitive and physically motivated parameters have definite units, but dedicated numerical techniques are necessary to form the respective vectors of regression data. Because in today's era of digital computers the creation of regressors can be performed effectively and reliably, in this study, we analyze the approach based on ordinary differential equations and related continuous-time models.

The paper is organized as follows. In Section 2 different techniques of numerical approximation of an original differential-equation model are presented. The discussion is focused on the method involving certain finite-horizon integration operators (Sagara filters). The classical least-squares procedure, and its asymptotic properties are briefly recalled in Section 3. Basic details, important for any practical implementation of the identification algorithms, which are insensitive to different noises or disturbances, are given in Section 4. In Section 5, the paper is summarized, and promising directions of further investigations are outlined.

2 Modeling of Continuous Systems

To maintain a physical interpretation of estimated parameters, a continuous-time model of the supervised process can be taken into account. For single-input single-output (SISO) systems an ordinary differential equation can be employed ($n > m \geq 0$):

$$y^{(n)}(t) + a_{n-1}y^{(n-1)}(t) + \mathrm{K} + a_0 y(t) = b_m u^{(m)}(t) + b_{m-1}u^{(m-1)}(t) + \mathrm{K} + b_0 u(t) \ , \tag{1}$$

where a_i and b_i stand for the unknown parameters. In general, the system (1) is subject to initial conditions: $u^{(m-1)}(0), \ldots, u(0), y^{(n-1)}(0), \ldots, y(0)$, which are responsible for the system's free response. For better readability of the further consideration, one can use the following transfer function (i.e. counterpart to the differential equation):

$$Y(s) = H(s) \cdot U(s) + F(s) \ , \tag{2}$$

$$H(s) = \frac{b_m s^m + b_{m-1}s^{m-1} + \mathrm{K} + b_0}{s^n + a_{n-1}s^{n-1} + \mathrm{K} + a_0} \ , \tag{3}$$

$$F(s) = \frac{d_{n-1}s^{n-1} + d_{n-2}s^{n-2} + \mathrm{K} + d_0}{s^n + a_{n-1}s^{n-1} + \mathrm{K} + a_0} \ , \tag{4}$$

$$d_i = y^{(n-i-1)}(0) + \sum_{j=i}^{n-2} a_{j+1}\, y^{(j-i)}(0) - \sum_{j=i}^{m-1} b_{j+1}\, u^{(j-i)}(0) \ , \tag{5}$$

where $i = 0 \ldots n-1$. The Laplace signal transforms $U(s)$ and $Y(s)$ correspond to the input $u(t)$ and output $y(t)$, respectively, $H(s)$ is the system transfer function, and $F(s)$ represents the transformed system's free response. Assuming that the modeled system (1) is stable, all roots in the denominator of (3) have negative real parts (the denominator polynomial is Hurwitz type). Then the system's free response decays asymptotically to zero, and the output $y(t)$ is bounded provided the input $u(t)$ is bounded.

Computer-aided identification of continuous systems relies on the recorded samples of the input and output signals. This approach, referred to as discrete identification of continuous systems, calls for effective methods of numerical approximation of the dynamics (1). Namely, the derivatives have to be replaced by certain discrete 'measures', enabling the 'new' model to maintain the original parameters.

In the simplest way, the 'discrete-continuous' delta (δ) operator [11] can be used to evaluate the consecutive derivatives represented in (1), and the differentials ($dx(t)$, $dx^{(2)}(t)$, …) of a considered signal $x(t)$ are replaced by the respective finite differences ($\Delta x(kT) = x(kT+T) - x(kT)$, $\Delta^2 x(kT) = x(kT+2T) - 2x(kT+T) + x(kT)$, …), for a given sampling time T standing for the time differentiate dt. The delta method, however, suffers from several drawbacks. With a high-pass nature of δ, for instance, additive noises that corrupt the measurement signals are amplified. Additionally, delta differentiation of discontinuous input signals (e.g. square waves) becomes problematic. What is more, the discretized model formed with the aid of this non-causal operator cannot be identified on-line (future data $x(kT+T)$, $x(kT+2T)$, … , must be known to evaluate the derivatives of $x(t)$ at the given sampling instant $t = kT$).

A solution made more robust to additive noises utilizes a low-pass operator of multiple integration (i.e. $1/s^n$) to transform the model (1). By performing this integration, equally on both sides of the original differential equation, an integral model (without derivatives) is obtained, leading (via numerical integration) to a discrete representation of (1). Unfortunately, this approach also brings about some implemental problems. First, the integral components themselves tend to infinity, even in the case of measurable (bounded) input-output signals. Second, the initial conditions of (1) cannot be disregarded, as the n-times integrated free response does not decay to zero. As a consequence, the initial conditions have to be included in the identified integral model, which increases the complexity of modeling and computations.

A promising solution that overcomes the above-mentioned problems was proposed by Sagara [10], where a finite-horizon integration filter is used

$$J^n x(t) = \int_{t-h}^{t} \int_{t_1-h}^{t_1} \Lambda \int_{t_{n-1}-h}^{t_{n-1}} x(t_n)\, dt_n\, dt_{n-1} \ldots dt_1 \tag{6}$$

to rearrange the differential equation. Note that (6) describes the multiple nth order integration of a signal over the time interval of a fixed length (h).

Discrete realization of (6) is straightforward by employing the methods of numerical integration. With a convenient rule of trapezoidal integration, for instance, the operation (6) subject to the rth derivative of a signal can be implemented as

$$I_r^n(q^{-1}) = Q_r(q^{-1})\,(1-q^{-1})^r\,(1+q^{-1}+\ldots+q^{-L+1})^n\ , \tag{7}$$

with

$$Q_r(q^{-1}) = (\tfrac{T}{2})^{n-r}\,(1+q^{-1})^{n-r}\ , \tag{8}$$

where q^{-1} is the delay operator ($q^{-1}\cdot x(kT) = x(kT-T)$), and the integration horizon (h) is expressed as a multiplicity (L) of the sampling time (i.e. $h = L\cdot T$).

Keeping in mind that simple procedures of discrete integration tend to accumulate numerical errors, an improvement was proposed by Kowalczuk [2]. With improved integration methods (i.e. splines) used in this processing, a robust version of the Sagara filter (6) can be directly obtained based on the following polynomial $Q_r(q^{-1})$:

$$Q_r(q^{-1}) = \tfrac{T^{n-r}}{(n-r+1)!}\,N_{n-r+1}(q^{-1})\ , \tag{9}$$

where the so-called normal polynomials ($N_0(q^{-1}) = N_1(q^{-1}) = 1,\ N_2(q^{-1}) = 1 + q^{-1},\ \ldots$) result from:

$$N_p(q^{-1}) = \sum_{i=1}^{p} \eta_{i,p}\ q^{-i+1}\ ,\ \eta_{i,p} = \sum_{j=1}^{i} (-1)^{i-j}\,\tbinom{p+1}{i-j}\,j^{\,p}\ . \tag{10}$$

Ultimately, the transformed model can be written down in a common regression form:

$$\gamma(k) = I_n^n y(k) = \boldsymbol{\varphi}^{\mathrm{T}}(k)\,\boldsymbol{\theta} + e(k)\ , \tag{11}$$

$$\boldsymbol{\varphi}(k) = [\,-I_{n-1}^{n} y(k) \ldots -I_0^n y(k)\quad I_m^n u(k) \ldots I_0^n u(k)\,]^{\mathrm{T}}\ , \tag{12}$$

$$\boldsymbol{\theta} = [\,a_{n-1} \ldots a_0\quad b_m \ldots b_0\,]^{\mathrm{T}}\ , \tag{13}$$

with index k representing the sampling instant $k\cdot T$, $\gamma(k)$ being a reference signal, $e(k)$ denoting an equation error, and $\boldsymbol{\varphi}(k)$ and $\boldsymbol{\theta}$ standing for the regression data and model parameters, respectively. There are several evident benefits of the presented modeling strategy: (i) The regressors (12), obtained as the result of FIR filtering (7) of measured input-output signals, are bounded. (ii) System initial conditions can be disregarded in (11), since the influence of the filtered free response is entirely eliminated after the elapse of time $n\cdot L\cdot T$. (iii) For the integration horizon ($h = L\cdot T$) tuned so that the normalized magnitude characteristics of the Sagara filter ($\sin(0.5\cdot\omega h)/(0.5\cdot\omega h)$) and of the identified system ($|(a_0/b_0)\cdot H(j\omega)|$) are closely matched, one obtains an efficient elimination of additive noises. (*iv*) As the regression model (11)–(13) retains the original parameterization of (1), any well-established estimation scheme can be used to estimate the original system parameters. Implementations of different identification algorithms, and their pertinent properties, are presented in the next section.

3 Classical Least-Squares Method

In the simplest approach, the estimates of the regression model (1) can be obtained by a classical method of least-squares (LS), wherein the identification algorithm is based on minimizing of the following quadratic index [8]:

$$V_{\mathrm{LS}}(\boldsymbol{\theta}) = \sum_{l=1}^{k} [\, e(l) \,]^2 = \sum_{l=1}^{k} [\, \gamma(l) - \boldsymbol{\varphi}^{\mathrm{T}}(l)\, \boldsymbol{\theta} \,]^2 \quad . \tag{14}$$

By setting the gradient of (14) to zero, one gets the LS estimator in an algebraic form

$$\hat{\boldsymbol{\theta}}(k) = \left[\sum_{l=1}^{k} \boldsymbol{\varphi}(l)\, \boldsymbol{\varphi}^{\mathrm{T}}(l) \right]^{-1} \left[\sum_{l=1}^{k} \boldsymbol{\varphi}(l)\, \gamma(l) \right] . \tag{15}$$

The above formula is mathematically simple, but its practical implementation suffers from cumbersome matrix inversion at each sampling instant. In order to solve this problem, the formulas in brackets are recursively prescribed, and the well-known "matrix inversion lemma" is used. As a result, the algorithm LS obtains the following (convenient) recursive representation:

$$\mathbf{P}(k) = \mathbf{P}(k-1) - \frac{\mathbf{P}(k-1)\, \boldsymbol{\varphi}(k)\, \boldsymbol{\varphi}^{\mathrm{T}}(k)\, \mathbf{P}(k-1)}{1 + \boldsymbol{\varphi}^{\mathrm{T}}(k)\, \mathbf{P}(k-1)\, \boldsymbol{\varphi}(k)} \quad , \tag{16}$$

$$\hat{\boldsymbol{\theta}}(k) = \hat{\boldsymbol{\theta}}(k-1) + \mathbf{P}(k)\, \boldsymbol{\varphi}(k) [\, \gamma(k) - \boldsymbol{\varphi}^{\mathrm{T}}(k) \hat{\boldsymbol{\theta}}(k-1) \,] \quad . \tag{17}$$

It suffices, for implemental reasons, that the covariance matrix $\mathbf{P}(k)$ at the start-up of the algorithm gets a huge diagonal, e.g.: $\mathbf{P}(0) = \mathrm{diag}\,(10^5, \ldots, 10^5)$.

The asymptotic behavior of the routine LS can be concluded based on the algebraic LS formula. By substituting (11) into (15) one acquires

$$\hat{\boldsymbol{\theta}}(k) - \boldsymbol{\theta} = \left[\frac{1}{k} \sum_{l=1}^{k} \boldsymbol{\varphi}(l)\, \boldsymbol{\varphi}^{\mathrm{T}}(l) \right]^{-1} \left[\frac{1}{k} \sum_{l=1}^{k} \boldsymbol{\varphi}(l)\, e(l) \right]_{k \to \infty} = \{ E[\boldsymbol{\varphi}(k)\, \boldsymbol{\varphi}^{\mathrm{T}}(k)] \}^{-1} E[\boldsymbol{\varphi}(k)\, e(k)] \tag{18}$$

where, for stationary ergodic processes, the averaging in time can be approximated by probabilistic measures (i.e. correlation functions). It is evident from (18) that the LS estimates are asymptotically consistent, if the residual error is uncorrelated with the regression data ($E[\boldsymbol{\varphi}(k)\, e(k)] = 0$). This takes place, for instance, when the process $e(k)$ is a sequence of zero-mean independent random variables (white noise). Unfortunately, in most cases, the residuals are correlated, and their corresponding mean value is not necessarily zero. What is more, the parameters of the identified system (1) can be time-varying, and different perturbations (like destructive outliers) can appear in measurement data. In order to challenge these issues, specific techniques and hints are subsequently put into practice.

4 Practical Implementation of Robust Estimators

In this section a practical implementation of the estimation procedure insensitive to noises, disturbances, and other perturbations, is discussed in detail.

Dedicated solutions are proposed as a remedy to the problems of tracking time-varying system parameters, and to the inconsistency of estimates invoked by correlated noises and outliers in the measurement data. Finally, conditions imposed on the identifiability of the system are formulated, and suggestions concerning the selection of persistently exciting inputs are given.

4.1 Tracking the Time-Varying Parameters

In the case of non-stationary systems, when the coefficients a_i and b_i presented in the differential equation evolve in time, the classical LS procedures (15) or (16)–(17) cannot be directly applied. This is so, because in the basic algorithm LS all measurement data is assumed to represent a common dynamics. In case of non-stationary systems, however, only actual data should be considered in the current evaluation of the identified parameters. The problem can be effectively overcome by re-defining the quality index (14), so that the impact of old data is much reduced

$$V_{\mathrm{LS}}(\boldsymbol{\theta}) = \sum_{l=1}^{k} \lambda^{k-l} \, [\, e(l) \,]^2 = \sum_{l=1}^{k} \lambda^{k-l} \, [\, \gamma(l) - \boldsymbol{\varphi}^{\mathrm{T}}(l)\, \boldsymbol{\theta} \,]^2 \quad . \tag{19}$$

The parameter λ (called the forgetting factor) used in the utility weighting mechanism ($w_l = \lambda^{k-l}$) usually falls within the range of $[0.9\,, 1]$. Minimization of the modified criterion (19) leads to the following algebraic solution:

$$\hat{\boldsymbol{\theta}}(k) = \left[\sum_{l=1}^{k} \lambda^{k-l} \, \boldsymbol{\varphi}(l) \boldsymbol{\varphi}^{\mathrm{T}}(l) \right]^{-1} \left[\sum_{l=1}^{k} \lambda^{k-l} \, \boldsymbol{\varphi}(l) \gamma(l) \right] \tag{20}$$

and its recursive counterpart

$$\mathbf{P}(k) = \frac{1}{\lambda} \left[\mathbf{P}(k-1) - \frac{\mathbf{P}(k-1)\, \boldsymbol{\varphi}(k)\, \boldsymbol{\varphi}^{\mathrm{T}}(k)\, \mathbf{P}(k-1)}{\lambda + \boldsymbol{\varphi}^{\mathrm{T}}(k)\, \mathbf{P}(k-1)\, \boldsymbol{\varphi}(k)} \right] , \tag{21}$$

$$\hat{\boldsymbol{\theta}}(k) = \hat{\boldsymbol{\theta}}(k-1) + \mathbf{P}(k) \boldsymbol{\varphi}(k) \, [\, \gamma(k) - \boldsymbol{\varphi}^{\mathrm{T}}(k) \hat{\boldsymbol{\theta}}(k-1) \,] \, , \tag{22}$$

both being realizations of the exponentially weighted least-squares routine (EW-LS). The tracking ability of the above algorithm depends on the length of the estimator's memory, or the effective memory of the estimator equal to $\Gamma = 1/(1-\lambda)$. Clearly, the non-weighted estimator (21)–(22) can be obtained from (16)–(17) by substituting $\lambda = 1$, which means that the basic LS algorithm has infinite memory ($\Gamma = \infty$).

4.2 Suppressing the Bias Influenced by Additive Noises

In practical situations, the observed analog signals are corrupted with high frequency additive noise. Such disturbances can be almost entirely eliminated using tuned low-pass filters. Unfortunately, basic discretization of the model (necessary in discrete-time processing) introduces additive noise to sampled data. Note that the measurement 'round-off' noise is a typical phenomenon resulting from the finite resolution of AD converters. Thus, a zero-mean disturbance sequence with variance dependent on the resolution of such a converter contaminates the sampled (discrete-time) signal.

Suppose now that the sampled output signal $y(k)$ is corrupted with an additive white noise $v(k)$. Then the numerical mechanization of the regression data yields

$$I_i^n\,[\,y(k)+v(k)\,]=I_i^n\,y(k)+I_i^n\,v(k) \tag{23}$$

for all components ($i=0\ldots n$) in the regression vector (12). As a result, the residual $e(k)$ represented in the regression model (11) takes the form of a moving average

$$e(k)=I_n^n\,v(k)+a_{n-1}\,I_{n-1}^n\,v(k)+\mathrm{K}\,+a_0\,I_0^n\,v(k)\ . \tag{24}$$

It is evident from (24) that with the correlated disturbance the LS estimates are certain to be asymptotically biased ($E\,[\boldsymbol{\varphi}(k)\,e(k)]\neq 0$). The consistency of the estimates can be improved by employing the idea of an instrumental variable (IV).

Assuming that such an instrumental (deterministic) vector $\boldsymbol{\xi}(k)$ is used in place of the original regression vector, one acquires the following consistency formula:

$$\hat{\boldsymbol{\theta}}(k)-\boldsymbol{\theta}=\left[\frac{1}{k}\sum_{l=1}^{k}\boldsymbol{\xi}(l)\,\boldsymbol{\varphi}^{\mathrm{T}}(l)\right]^{-1}\left[\frac{1}{k}\sum_{l=1}^{k}\boldsymbol{\xi}(l)\,e(l)\right]_{k\to\infty}=\{E[\boldsymbol{\xi}(k)\,\boldsymbol{\varphi}^{\mathrm{T}}(k)]\}^{-1}E[\boldsymbol{\xi}(k)\,e(k)]\ . \tag{25}$$

With the instrument uncorrelated with the process $e(k)$ ($E\,[\boldsymbol{\xi}(k)\,e(k)]=0$)), the asymptotic behavior of the estimator is much improved. It is evident from the above reasoning that the procedure IV can be derived from the LS routine by replacing $\boldsymbol{\varphi}(k)$ with $\boldsymbol{\xi}(k)$ (while the row vector $\boldsymbol{\varphi}^{\mathrm{T}}(k)$ remains unmodified). This gives instantly the following algebraic form

$$\hat{\boldsymbol{\theta}}(k)=\left[\sum_{l=1}^{k}\lambda^{k-l}\,\boldsymbol{\xi}(l)\,\boldsymbol{\varphi}^{\mathrm{T}}(l)\right]^{-1}\left[\sum_{l=1}^{k}\lambda^{k-l}\,\boldsymbol{\xi}(l)\,\gamma(l)\right] \tag{26}$$

and its recursive implementation

$$\mathbf{P}(k)=\frac{1}{\lambda}\left[\mathbf{P}(k-1)-\frac{\mathbf{P}(k-1)\,\boldsymbol{\xi}(k)\,\boldsymbol{\varphi}^{\mathrm{T}}(k)\,\mathbf{P}(k-1)}{\lambda+\boldsymbol{\varphi}^{\mathrm{T}}(k)\,\mathbf{P}(k-1)\,\boldsymbol{\xi}(k)}\right], \tag{27}$$

$$\hat{\boldsymbol{\theta}}(k)=\hat{\boldsymbol{\theta}}(k-1)+\mathbf{P}(k)\boldsymbol{\xi}(k)\,[\,\gamma(k)-\boldsymbol{\varphi}^{\mathrm{T}}(k)\hat{\boldsymbol{\theta}}(k-1)\,] \tag{28}$$

of the intended instrumental variable routine (with the auxiliary weighting mechanism using the forgetting factor λ). Commonly, a running-in-parallel procedure LS is used to support the recursive IV algorithm at the start-up phase. Eventually, for a suitably generated instrumental variable $\boldsymbol{\xi}(k)$, the processing is continued using the proper formulas (27) – (28).

Different manifestations of instrumental variables can be found in the literature [12]. In the case of continuous-time systems, however, a general method employing a noise-free evaluation of the output process $y(k)$ can be applied [10]. In a practical solution the input $u(k)$ is assumed to be deterministic, while a noise-free measure of $y(k)$ results from the following auxiliary filtering (where $a_n = 1$):

$$\hat{y}(k) = \frac{\sum_{i=0}^{m} \hat{b}_{m-i} \, [\, 2(1-q^{-1}) \,]^{m-i} \, [\, T(1+q^{-1}) \,]^{n-m+i}}{\sum_{i=0}^{n} \hat{a}_{n-i} \, [\, 2(1-q^{-1}) \,]^{n-i} \, [\, T(1+q^{-1}) \,]^{i}} \, u(k) \quad . \tag{29}$$

Clearly, the estimated system transfer function (2) involved in processing (29) is accommodated to the discrete domain by using the well-known bilinear (Tustin's) operator: $s^{-1} \leftrightarrow (T/2) \cdot (1+q^{-1}) / (1-q^{-1})$. Hence, the instrumental variable $\boldsymbol{\xi}(k)$ is [10]:

$$\boldsymbol{\xi}(k) = [\, -I^n_{n-1}\hat{y}(k) \ldots -I^n_0\hat{y}(k) \quad I^n_m u(k) \ldots I^n_0 u(k) \,]^{\mathrm{T}} \quad , \tag{30}$$

where the noise-free output (29) substitutes the original measurement $y(k)$ contaminated by noise $v(k)$. Naturally, the variable (30), involving the deterministic quantities $u(k)$ and (29), is uncorrelated with the residual $e(k)$.

It is worth noticing that the method IV performs properly provided the residual $e(k)$ is a zero-mean process. What is more, neither correlation in the regression data, nor the probability distribution of $e(k)$, can disturb the consistency of the IV estimates.

4.3 Elimination of DC Offsets

Usually, it is assumed that the equation error $e(k)$ is a zero-mean process. It may happen, however, that a DC offset appears in the residual signals (disturbances) and therefore affects the accuracy of the estimates. In order to challenge this problem the regression model (11) can be generalized so as to embrace such offsets. In such a solution, the residual $e(k)$ is assumed to be the sum of an offset $\bar{e}$ and a zero-mean process $\tilde{e}(k)$. Thus, the extended model takes the common form (11) with

$$\boldsymbol{\varphi}(k) = [\, -I^n_{n-1}y(k) \ldots -I^n_0 y(k) \quad I^n_m u(k) \ldots I^n_0 u(k) \quad 1 \,]^{\mathrm{T}} \quad , \tag{31}$$

$$\boldsymbol{\theta} = [\, a_{n-1} \ldots a_0 \quad b_m \ldots b_0 \quad \bar{e} \,]^{\mathrm{T}} \quad . \tag{32}$$

Note that by applying the method EW-LS, the tracking of non-stationary parameters along with a DC offset varying in time, can be effectively implemented.

4.4 Systems Identification Insensitive to Outliers

Efficient elimination of outliers in measurements is probably the most challenging issue in the area of robust identification. It is well-known that the classical method LS (derived from minimization of a quadratic criterion) is sensitive to such errors in data. Contrary to this, an estimator in the sense of the least sum of absolute errors (LA) usually generates reliable estimates irrespective of sporadic outliers appearing in the recorded input-output signals. The procedure of least-absolute errors (LA) results from minimization of the following index [1, 4]:

$$V_{\mathrm{LA}}(\boldsymbol{\theta}) = \sum_{l=1}^{k} \lambda^{k-l} \left| e(l) \right| = \sum_{l=1}^{k} \lambda^{k-l} \frac{[\, \gamma(l) - \boldsymbol{\varphi}^{\mathrm{T}}(l)\, \boldsymbol{\theta} \,]^2}{\left| e(l) \right|} \ , \tag{33}$$

where, as before, the weighting mechanism ($\lambda \le 1$) can be used for tracking variable parameters. As the quality function cannot be minimized analytically, an iterative procedure ($p = 0, 1, \dots$) of successive approximations can be put into practice [5]:

$$\hat{e}^{\,/p/}(l) = \gamma(l) - \boldsymbol{\varphi}^{\mathrm{T}}(l)\, \hat{\boldsymbol{\theta}}^{\,/p/} \ , \tag{34}$$

$$\hat{\boldsymbol{\theta}}^{\,/p+1/} \approx \left[\sum_{l=1}^{k} \lambda^{k-l} \frac{\boldsymbol{\varphi}(l)\boldsymbol{\varphi}^{\mathrm{T}}(l)}{\left| \hat{e}^{\,/p/}(l) \right|} \right]^{-1} \left[\sum_{l=1}^{k} \lambda^{k-l} \frac{\boldsymbol{\varphi}(l)\gamma(l)}{\left| \hat{e}^{\,/p/}(l) \right|} \right] , \tag{35}$$

where the LS estimate (20) of $\boldsymbol{\theta}$ is used at the start-up ($p = 0$) of this iterative scheme. Note that estimates $\hat{e}(l)$ of the model errors $e(l)$ are based on current estimates of $\boldsymbol{\theta}$. It can be proved that the index (33) calculated for a sequence of iteratively obtained estimates (35), is decreasing [5]. With this, the following condition (with a user-defined threshold $\Delta_{\min}$) that terminates the iterative processing can be formulated as

$$\left| V_{\mathrm{LA}}(\hat{\boldsymbol{\theta}}^{\,/p+1/}) - V_{\mathrm{LA}}(\hat{\boldsymbol{\theta}}^{\,/p/}) \right| < \Delta_{\min} \ . \tag{36}$$

It should be taken into account that the iterative routine suffers from a numerical problem of small divisors in (35). This can be overcome by substituting close-to-zero estimates (34) of errors with an assumed boundary value ($\varepsilon_{\min} > 0$).

An approximate recursive realization of the method LA can also be obtained. What is more, the concept of instrumental variables can be applied in an approximate LA realization. Based on similar-as-before rearrangements of the expressions in the brackets of (35), and by applying the "matrix inversion lemma", the resulting routines (involving a single iteration only) assume the following form:

$$\mathbf{P}(k) = \frac{1}{\lambda} \left[\mathbf{P}(k-1) - \frac{\mathbf{P}(k-1)\, \boldsymbol{\psi}(k)\, \boldsymbol{\varphi}^{\mathrm{T}}(k)\, \mathbf{P}(k-1)}{\lambda \left| \varepsilon(k) \right| + \boldsymbol{\varphi}^{\mathrm{T}}(k)\, \mathbf{P}(k-1)\, \boldsymbol{\psi}(k)} \right] , \tag{37}$$

$$\hat{\boldsymbol{\theta}}(k) = \hat{\boldsymbol{\theta}}(k-1) + \mathbf{P}(k)\, \boldsymbol{\psi}(k)\, \mathrm{sgn}[\, \varepsilon(k) \,] \ , \tag{38}$$

where $\varepsilon(k) = \gamma(k) - \boldsymbol{\varphi}^{\mathrm{T}}(k)\,\hat{\boldsymbol{\theta}}(k-1)$. With $\boldsymbol{\psi}(k) = \boldsymbol{\varphi}(k)$ one has the scheme LA, and $\boldsymbol{\psi}(k) = \boldsymbol{\xi}(k)$ leads to the routine $\mathrm{IV}_{\mathrm{abs}}$. Note that generation of the variable $\boldsymbol{\xi}(k)$ can be identical as in the classical estimator. A numerically convenient initiation of the recursive scheme utilizes the LS (or IV) estimates at the start-up of identification. After starting-up of iterations, the LS (IV) assistance is switched off, and the processing is continued with the proper routine LA ($\mathrm{IV}_{\mathrm{abs}}$).

The analysis of the asymptotic behavior of the above approximate algorithms LA and $\mathrm{IV}_{\mathrm{abs}}$ resembles the respective considerations given in Section 4.2. For a suitably large k (i.e. in the limiting case of $k \to \infty$) the averaging in time can be approximated by probabilistic measures (ergodicity). Thus the estimates can be expressed as

$$\hat{\boldsymbol{\theta}}(k) - \boldsymbol{\theta} = \left[\frac{1}{k}\sum_{l=1}^{k}\boldsymbol{\psi}(l)\,\boldsymbol{\varphi}^{\mathrm{T}}(l)\right]^{-1}\left[\frac{1}{k}\sum_{l=1}^{k}\boldsymbol{\psi}(l)\,e(l)\right]_{k\to\infty} = \{E[\boldsymbol{\psi}(k)\boldsymbol{\varphi}^{\mathrm{T}}(k)]\}^{-1}E[\boldsymbol{\psi}(k)e(k)] \quad (39)$$

for the LA ($\boldsymbol{\psi}(k) = \boldsymbol{\varphi}(k)$) and $\mathrm{IV}_{\mathrm{abs}}$ ($\boldsymbol{\psi}(k) = \boldsymbol{\xi}(k)$) methods, respectively. Assuming, for instance, that the residual $e(k)$ is a zero-mean white noise sequence, the process $\operatorname{sgn}[e(k)]$ retains the white noise properties (the deterministic 'sign' function of the error modifies the probability distribution of the error only). Therefore, the respective correlation equals zero ($E\{\boldsymbol{\varphi}(k)\operatorname{sgn}[e(k)]\} = 0$), and the LA estimates are certain to be bias-free. In the case of the errors correlated with the regression data, in turn, the bias problem is overcome by introducing the instrumental variable $\boldsymbol{\xi}(k)$. Since this deterministic instrument is neither correlated with $e(k)$, nor correlated with $\operatorname{sgn}[e(k)]$, the result $E\{\boldsymbol{\xi}(k)\operatorname{sgn}[e(k)]\} = 0$ verifies the asymptotic consistency of the considered $\mathrm{IV}_{\mathrm{abs}}$ estimates.

4.5 Other Design Issues

Among other design issues, the selection of a persistently exciting input $u(k)$ needs to attract some attention. In order to guarantee the identifiability of the observed continuous-time system (1), certain conditions must be satisfied by the excitation signal used for identification. Loosely speaking, the input $u(k)$ should be sufficiently exciting, so as to extract full information from the examined process. Taking into consideration the system's frequency characteristics, the persistent excitation means that the spectral density of the process $u(k)$ is non-zero in at least p points. It is intuitively evident that a sufficiently exciting input should be persistent enough to activate all the modes of the identified system.

There are many stochastic excitations: the white noise process and PRBS (pseudo random binary sequence) are the most representative. Employment of these signals requires, however, that a reliable pseudo-random number generator is available. Therefore, deterministic signals of a sufficient order can also be applied in practical identification procedures. For example, a rich selection of several harmonic signals is usually sufficient. By applying such periodic signals, the problem of the undesired DC offsets is overcome, and the noise-to-signal ratio is improved. Having in mind that a single sinusoid ($\sin \omega k$) is persistently exciting of order 2, a practical rule for the Sagara models (11) – (13) of order n can be put into practice [10]: "It suffices that the input $u(k)$ is represented by a sum of at least $n+1$ sinusoids with different (normalized) frequencies (ω) placed within the range of $(0, \pi)$". It should be assured, however, that the frequencies contained in a multi-harmonic excitation do not coincide with the zeros of the FIR integrating filters (7).

5 Conclusions

In this study the methods of identification insensitive to perturbations (like noises, outliers, and DC offsets) have been discussed. First, a practical method of discrete modeling of continuous systems has been explained. With the aid of dedicated finite-horizon (FIR) integrating filters a discrete regression model preserving the original system parameterization can be obtained. It is of major importance that the regressors of the resultant model are numerically 'stable', while the system free response (induced by non-zero initial conditions) does not influence the modeling accuracy.

Next, the well-known algorithm LS and its basic properties have been recalled. In order to improve the quality of parameter estimation, substantial modifications of the basic procedure need to be put into practice. Namely, tracking the time-varying system parameters can be implemented by means of a weighting mechanism. A radical suppression of the asymptotic bias of estimates, in turn, results from the employment of instrumental variables. Yet, with the necessarily extended regression model, undesired DC offsets in additive disturbances become irrelevant. Finally, the challenging problem of estimates heavily influenced by outliers in data, is effectively overcome by using an estimator in the sense of the least sum of absolute errors. In conclusion, practical suggestions concerning the selection of proper excitations are given.

Also, a brief outline of some promising directions of further investigations in the area of identification of continuous dynamics can be given as follows:

1. Identification of models with input delay: In this case the quality criterion being non-linear with regard to the estimated time delay creates the fundamental difficulty. Some practical solutions, allowing for on-line identification of both the delay and system parameters, have been reported by Kowalczuk and Kozłowski [6, 7].
2. Identification of distributed parameter systems: The main problem is attributed to handling the models taking the form of partial differential equations. Using an extended Sagara filter (with finite-horizon integration in time and in space), however, a discrete-time counterpart model with the retained parameters follows at once [9].

3. Identification of non-linear models: Existing methods are often dedicated for practical industrial applications. In the literature, the Hartley modulating functions subject to Hammerstein models are usually considered. For some models (differential equations with non-linear static parts, for instance), a dedicated solution involving a simple Taylor's approximation can be sufficiently effective [5].

It is also worth noticing that some unconventional methods, such as artificial intelligence approaches, including genetic algorithms, can also be successfully applied in systems implementing dedicated identification routines.

References

1. Janiszowski, K.: To estimation in the sense of the least sum of absolute errors. In: 5th International Symposium on Methods and Models in Automation and Robotics, Międzyzdroje, Poland, vol. 2, pp. 583–588 (1998)
2. Kowalczuk, Z.: Discrete approximation of continuous-time systems – a survey. In: IEE Proceedings-G, Circuits, Devices and Systems, vol. 140, pp. 264–278 (1993)
3. Kowalczuk, Z., Kozłowski, J.: Continuous-time approaches to identification of continuous-time systems. Automatica **36**, 1229–1236 (2000). doi: 10.1016/S0005-1098(00)00033-9
4. Kowalczuk, Z., Kozłowski, J.: Non-quadratic quality criteria in parameter estimation of continuous-time models. IET Control Theory and Applications **5**, 1494–1508 (2011). doi: 10.1049/iet-cta.2010.0310
5. Kozłowski, J., Kowalczuk, Z.: Robust to measurement faults, parameter estimation algorithms in problems of systems diagnostics (In Polish: Odporne na przekłamania pomiarowe algorytmy estymacji parametrycznej w zagadnieniach diagnostyki systemów). In: Kowalczuk, Z., and Wiszniewski, B. (eds.) Intelligent information extraction for diagnostic purposes, pp. 221–240. Pomeranian Science and Technology Publishers, Gdańsk (2007)
6. Kozłowski, J., Kowalczuk, Z.: Continuous-time delay system identification insensitive to measurement faults. In Kowalczuk, Z. (ed.) Diagnosis of Processes and Systems, pp. 177–184. Pomeranian Science and Technology Publishers, Gdańsk (2009)
7. Kozłowski, J., Kowalczuk, Z.: On-line parameter and delay estimation of continuous-time dynamic systems. Int. J. Appl. Math. Comput. Sci. **25**, 223–232 (2015). doi: 10.1515/amcs-2015-0017
8. Ljung, L.: System identification. Theory for the user. Prentice-Hall Inc., Englewood Cliffs, New Jersey, USA (1987)
9. Sagara, S., Zhao, Z.Y.: Identification of system parameters in distributed parameter systems. In: 11th IFAC Triennial World Congress, Tallin, Estonia, pp. 471–476 (1990)
10. Sagara, S., Zhao, Z.Y.: Numerical integration approach to on-line identification of continuous-time systems. Automatica **26**, 63–74 (1990). doi: 10.1016/0005-1098(90)90158-E
11. Söderström, T., Fan, H., Carlsson, B., Bigi, S.: Least squares parameter estimation of continuous-time ARX models from discrete-time data. IEEE Trans. On Automatic Control **42**, 659–673 (1997). doi: 10.1109/9.580871
12. Söderström, T., Stoica, P.: Comparison of some instrumental variable methods – consistency and accuracy aspects. Automatica **17**, 101–115 (1981). doi: 10.1016/0005-1098(81)90087-X