Synthesis of smoothing cubic spline in non-parametric identification technical systems’ algorithm

Y Voskoboinikov1,* and V Boeva1

1Novosibirsk State University of Architecture and Civil Engineering (Sibstrin), 113, Leningradskaya str., Novosibirsk, 630008, Russia
E-mail: * voskob@mail.ru

Abstract. Many technical systems are described by mathematical models in the form of an integral Voltaire equation of the first kind with a difference kernel. The nonparametric identification problem for such systems will be reduced to constructing an estimate of the impulse transition function of the identified system based on the measured (with noise) the input and output signals’ values. A rectangular signal of constant amplitude is applied to the input of the identified system at some point in time in some identification schemes that are widely used in practice. For such an input signal, the pulse transition function is defined as the first derivative of the system output signal. However, the derivative calculation belongs to the class of incorrectly posed problems, and an essential feature of this problem is the calculated derivative instability to the errors in recording the output signal. For stable computation, various algorithms for smoothing experimental data are used. The most effective of them are smoothing cubic splines. Boundary conditions are specified to uniquely determine the coefficients of these splines. Unfortunately, traditional natural boundary conditions (zero second derivatives of the spline) do not allow taking into account the features of the identification problem. Therefore, we propose a smoothing algorithm, based on smoothing cubic splines, which allows one taking into account information about the identifiable impulse transition function in sufficient (to increase the accuracy of identification) in this research paper. The studies show the effectiveness of the proposed smoothing algorithm and the entire identification procedure as a whole.

1. Introduction
Integrated models are often used to model various dynamic systems [1,2]. If the system parameters do not change in time, then such a system is called stationary and its behavior is described by the Voltaire integral equation of the first kind with a difference kernel [1]:

\[ \int_0^t k(t-\tau) \phi(\tau) d\tau = f(t), \quad t \in [0,T] \] (1)

Here \( k(t) \) is impulse transition function (ITF) of the system; \( \phi(\tau), f(t) \) – its input and output signals. For a physically realizable system, the condition \( k(t)=0 \) at \( t < 0 \).

The nonparametric identification problem of systems (1) is to calculate the estimate for ITF \( k(t) \) from the measured values of the input and output signals of the identified system. Such a problem belongs to the class of incorrectly posed problems [3] and special regularizing algorithms are used in order to solve it (for example, [4]).

In practice, in order to identify stationary systems, a circuit is used often (because of its simplicity) where a step signal is applied to the input of the identified system (at the moment \( t=0 \), which
amplitude is constant. We should note that, if the amplitude is 1, then such a signal is called the Heaviside function (fulfillment of this condition is assumed in the future). For such an input signal, it can be shown [1] that:

\[ k(t) = \frac{d}{dt}f(t) \]

(2)

Here \( f(t) \) is output signal (system response), if the Heaviside function is applied to the input. Despite the well-developed algorithms for numerical differentiation, the use of (2) in practice is associated with the same incorrectness problem as it is in the direct solution of the equation (1), since the differentiation operation is also an incorrectly posed problem [5,6]. Most often this is manifested in the instability of the differentiation operation, when even a small noise level of the output signal registration causes very large errors in the obtained (based on (2)) ITF estimate.

The recorded (with measurement errors) \( \tilde{f}(t) \) signal must initially be smoothed (i.e., filtered to one degree or another) from the measurement errors (before the differentiation operation) in order to obtain a stable estimate of ITF. For this purpose, various methods can be used: the finite difference calculation (with an appropriate choice of the time discretization step); local spatial filters [7,8]; wavelet filtering algorithms [9,10]; smoothing splines [11-13].

The most universal for practice are smoothing splines, and more specifically, smoothing cubic splines (SCS), which allow not only smoothing out noisy data, but also calculating the first derivative of a smoothed function \( \tilde{f}(t) \). It is necessary to set the values of the boundary conditions at points \( t_1=0, t_N=T \) for an unambiguous calculation of SCS coefficients [11]. As a rule, the so-called natural boundary conditions are used when zero values of the second spline derivatives are specified. Unfortunately, for the identification problem to be solved, setting such boundary conditions leads to large identification errors in the neighbourhoods of the extreme points \( t_1, t_N \).

- development of an algorithm for SCS constructing with boundary conditions adequate to the identification problem being solved, namely, setting the values of the first derivative at extreme points;
- development of the estimating statistical algorithm for the optimal value of the smoothing parameter from the condition of minimum root-mean-square error of smoothing.

Thus, the problem of synthesizing SCS for a specific identification algorithm (2) of a dynamic system is solved.

2. Mathematical and Methods

Possess that registered in nodes \( 0=t_1<t_2<...<t_N=T \) noisy output \( \tilde{f}(t) \) the form can be:

\[ \tilde{f}(t) = \eta(t) \]

(3)

Here \( \eta(t) \) is zero mean dispersion noise \( \sigma^2 \). To smooth out noisy values \( \tilde{f}(t) \) let’s turn to smoothing cubic splines (SCS). We should remind [7], that smoothing cubic spline \( S_f \) on each segment \( [t_i, t_{i+1}) \), \( i=1,..N-1 \) is the third degree polynomial of the form:

\[ S_{f,i}^{3}(t) = a_i + b_i(t-t_i) + c_i(t-t_i)^2 + d_i(t-t_i)^3 \]

(4)

and is twice continuously differentiable over the entire interval \( [0,T] \). Left and right boundary conditions are specified in order to unambiguously calculate SCS coefficients \( a_i, b_i, c_i, d_i \). For example, natural boundary conditions determine the zero values of the second derivatives at the nodes \( t_1, t_N \):

\[ S_{f,a}^{\prime\prime}(t) \bigg|_{t=t_1} = S_{f,a}^{\prime\prime}(t) \bigg|_{t=t_N} = 0 \]

(5)

It was shown [7] that SCS with natural boundary conditions delivers the minimum value to the functional:
\[ F_{\alpha}(S) = \alpha \int_{t_1}^{t_N} |S'(t)|^2 \, dt + \sum_{i=1}^{n} p_i^{-1}(\tilde{f}_i - S(t_i))^2 \]

where \( p_i^{-1}, i=1, \ldots, N \) are the weighting factors. It can be seen that the functional is determined by both the “smoothness” of the spline (the first term) and the “proximity” of the spline to the given values \( \{\tilde{f}_i\} \). Changing the smoothing parameter \( \alpha \) in the interval \([0, \infty)\), you can change the smoothing error. The smoothing parameter’s value that minimizes average square error (ASE) of smoothing is called the optimal parameter and denote \( \alpha_{\text{opt}} \). The estimation algorithm will be considered later.

Unfortunately, the specification of boundary conditions \( S_{t_0}^{(2)} \) can lead to significant identification errors (SCS differentiation errors) for small and large values of SCS argument. In order to illustrate this fact, Figure 1 shows the values of the “exact” ITF (solid curve) and the estimate \( \tilde{S}_{\alpha_{\text{opt}}}^{(2)} \) (dashed curve) calculated by differentiating the SCS, calculated under \( \alpha = \alpha_{\text{opt}} \) and boundary conditions (5), while the measurement noise had a relative level of 0.15.

**Figure 1.** Impulse transient function estimates.

Rather great identification error is visible at small values instead of zero at \( t = 0 \) evaluating \( \tilde{S}_{\alpha_{\text{opt}}}^{(2)} \), but the influence of the left boundary condition decreases as the argument increases. A similar error can be observed for time values close to the right end of the interval \( \tilde{S}_{\alpha_{\text{opt}}}^{(2)} \). It is necessary to specify boundary conditions based on the specifics of the problem being solved to reduce such errors. For example, for a number of dynamic systems (second-order vibrational links and others) it is known that \( k(0) = 0, k(T) = 0 \). It is advisable to set the boundary conditions with the values of the first derivative, given this and similar a priori information, which will undoubtedly lead to an increase in the identification accuracy. Therefore, we proceed to the algorithm construction for calculating the spline coefficients under such boundary conditions.

We compose a system of equations (using the results of [11]) to calculate the values \( S_{t_0}^{(2)} \) in the case, when the first derivatives’ values are set as boundary conditions on the left:

\[ s_i = S'(t_i), s_N = S'(t_N). \]

It is shown that such a system can be written in the following form:
The coefficients of the system are determined by the relations where $h_i$ are the steps between corresponding grid nodes.

After calculating of this system solution, the spline coefficients are determined by the expressions:

$$a_i = \frac{\hat{r}_{H_i}}{h_i} - \alpha p_i; \quad a_N = \frac{\hat{r}_{H_N}}{h_N} - \alpha p_N^{M_{N-1}M_N};$$

$$a_i = \frac{\hat{r}_{H_i}}{h_i} - \alpha p_i \left[ \frac{M_{i+1}M_i - M_iM_{i-1}}{h_i - h_{i-1}} \right], \quad i=2,3,...,N-1;$$

$$b_i = \frac{\alpha p_i}{h_i} + \frac{\alpha p_{i+1}}{h_{i+1}} - \frac{M_{i+1}M_i}{6h_i} + \frac{M_iM_{i-1}}{6h_{i-1}}, \quad i=1,2,...,N-1.$$
\[ \rho \frac{1}{n} \sum_{i=1}^{N} \tilde{H}_i \cdot e_i(\alpha) \]  

(8)

here \( e_i(\alpha) = \tilde{H}_i - \text{Si}_{f,\alpha} \) is residual of the \( i \)-th dimension. For wavelet filtering algorithms, it is proved that if \( \rho_W \) at a value \( \alpha \) satisfies the inequality

For wavelet filtering algorithms, it is proved that if \( \rho_W \), for some \( \alpha \) value, it satisfies the inequality

\[ v \frac{\beta}{2} N W \frac{1}{2N} \]  

(9)

such a value can be taken as an estimate for \( \alpha_{opt} \) (we denote this value as \( \alpha_W \)). In inequality (9), the quantities \( v \frac{\beta}{2} N W \frac{1}{2N} \) are quantiles of \( \chi^2 \) distribution with \( N \) degrees of the levels freedom, \( \frac{1}{2}, \frac{1}{2} \) respectively. The value \( \beta \) determines the probability of the first kind error, during the hypothesis testing about the optimality of the estimate \( \alpha_W \) and, \( \beta = 0.05 \) as a rule. We should highlight that \( \alpha_W \) calculation reduces to solving a nonlinear equation iterative algorithms.

\[ \rho_W(\alpha) = N \]  

(10)

As the next approximate \( \alpha_W \) solution of \( a^{(n)} \) equation (10) is accepted, this satisfies inequality (9). Since it is not necessary to find the exact solution to the nonlinear equation (10), but only a solution satisfying (9), one can use “slow” iterative procedures, for example, the dichotomy method.

3. Results

A large computational experiment was conducted in order to verify the operability and effectiveness of the proposed identification algorithm using boundary conditions in the form of the first derivative values. Let us stay only on the results of one studies’ series. The ITF of the second-order vibrational link (in Figure 1 is a solid curve) \( k(t) \) was taken. The values of the input and output signals were distorted by random normally distributed errors with relative levels \( \delta_i = \| \tilde{f}_i - f_i \| / f_i \), where \( \| \| \) is the Euclidean norm of the vector. The identification error was determined by the relative error \( \delta = \| \tilde{f}_{aw} \| / \| k \| \), where \( \tilde{f}_{aw} \) is the ITF estimate calculated by SCS. The measurements number of the output signal is \( N = 80 \).

Initially, we consider the studies’ results on the choice of the smoothing parameter, based on the given statistical criterion. In Figure 2, the solid curve shows the relative filtering error \( \delta_k \) using SCS for various parameter values \( \alpha \). We see the presence of a minimum at \( \alpha = 3_{opt} \). Dashed lines show the boundaries of inequality (9), and the dotted curve shows the values \( \rho_W \). As \( \alpha_W \) estimate, we take the quantities \( \alpha \), which \( \rho_W \) are located between dashed lines for. It can be seen that these values correspond to the minimum values of the filtering error.
Figure 2. The way of a smoothing parameter choosing.

A computational experiment also showed that a spline, constructed with $\alpha = \alpha_{opt}$ has:

- a smoothing error slightly (by 5-8%) exceeding the smoothing error with a parameter $\alpha = \alpha_{opt}$ (which can only be found in a computational experiment);
- the smoothing error is significantly (by 15-35%) less in comparison with other methods of parameter selection (for more details, see [10,15]).

All this allows us to conclude that it is advisable to use a smoothing spline with $\alpha = \alpha_{opt}$ for stable calculation of the input and output signals’ derivatives.

We point out on the value $\delta_k$ is random, since it depends on the specific implementation of the measurement noise. Therefore, the average value $\bar{\delta}_k$ was calculated (the mathematical expectation estimate of this random variable) for a sample of 30. Table 1 shows these average values.

| $\delta_f$ | $\bar{\delta}_k$ |
|------------|-----------------|
| 0.02       | 0.031           |
| 0.05       | 0.072           |
| 0.10       | 0.112           |
| 0.15       | 0.176           |

4. Conclusion

An analysis of the table data and other of computational experiments’ results with other ITF forms shows that, the proposed identification algorithm has a transmission coefficient of the initial data error to $\frac{\delta_k}{\delta_f}$ identification error not exceeding 2, even with a high noise level of the output signal measurement. This indicates a good stability of noise identification algorithm of the measurement output signal of an identifiable system. In addition, the use of SCS allows you to process measurements, obtained with an unequal sampling step in time, which is a significant limitation for other smoothing and differentiation methods. Therefore, there is every reason to recommend the proposed algorithm for solving practical problems of identifying stationary dynamic systems.

References

[1] Greblicki W and Pawiak M 2008, *Nonparametric System Identification*, Cambridge: Cambridge University Press, p 400

[2] Pillonetto G De and G Nicolao G New trends in nonparametric linear system identification
Modelling Methodology for Physiology and Medicine, Elsevier, 2, pp 111-129 (https://doi.org/10.1016/B978-0-12-411557-6.00006-92014)

[3] Neto F D M and Da. Silva D a 2013 An Introduction to Inverse Problems with Applications, Berlin: Springer, p 427

[4] Voskoboinikov Yu E and D.A. Krysov D A 2018 Local regularizing algorithm for nonparametric identification of an object with an inaccurate input signal, Scientific Bulletin of Nizhny Novgorod State Technical University, 70 (1) pp 13-20 (DOI: 10.17212 / 1814-1196-2018-1-19-38 2018)

[5] Ahn S, Choi U J and Ramm A G 2006 A scheme for stable numerical differentiation, J. Comp. Appl. Math. 186 pp 325-34

[6] Choi W Y 2009 A new method for stable numerical differentiation Curr. Appl. Phys.9 pp 1463-66

[7] Voskoboinikov Yu E 2019 Optimal parameter estimation of spatial-local signal filtering algorithms Optoelectronics, Instrumentation and Data Processing 55 (3) pp 222-29

[8] Voskoboinikov Yu E 2019 Optimal parameters estimation of spatial-local filters, Autometry, 3, pp 13-20 (DOI: 10.15372 / AUT20190302 2019)

[9] Gao R X 2011 Wavelets Theory and Applications for Manufacturing, New York Dordrecht Heidelberg London, Springer, p 241 (10.1007/978 1 4419 1545 0 2011)

[10] Voskoboinikov Yu E and Gochakov A V 2011 Estimation optimum threshold sizes in a wavelet filtration algorithms Optoelectronics, Instrumentation and Data Processing, 47 (2), pp 3-14

[11] Wang Y 2011 Smoothing Splines Methods and Applications (ser. Monographs on Statistics and Applied Probability), Chapman & Hall, 121 p 347

[12] Yang L and Hong Y 2017 Adaptive penalized splines for data smoothing Comp. Statistics & Data Analysis, 108 pp 70-83 (https://doi.org/10.1016/j.csda.2016.10.022

[13] Crambes C, Kneip A 2009 Smoothing splines estimators for functional linear regression The Annals of Statistics, 37 (1), pp 35-72 (DOI: 10.1214/07-AOS563 2009)

[14] Voskoboinikov Yu E 2017 Estimation of optimal parameters a class of multiplicative algorithms wavelet filtration of images Optoelectronics, Instrumentation and Data Processing, 53 (4), pp 13-21

[15] Lee T C M 2006 Smoothing parameter selection for smoothing splines: a simulation study Comp. Statistics & Data Analysis, 42 (1-2), pp 139-48 (DOI: 10.1016/S0167-9473(06)00159-7 2006)