SOME PROBLEMS OF PASSIVE EXPERIMENTAL DATA PROCESSING

Vladimir Troyanovskyi
Chair of computer science and the software of computing systems
Moscow Institute of Electronic Engineering
Russian Federation
troy40@mail.ru

Olga Serdyuk
Chair of computer science and the software of computing systems
Moscow Institute of Electronic Engineering
Russian Federation
olenkaalex@gmail.com

Abstract

This article illustrates the problems arising at data processing of passive experiment. Some specifics of passive experiment and subtleties of data processing with aid of available mathematical methods are considered. It is shown how the parity of the sampling period, time parameters of signal's properties and dynamic properties of object in common influence on the researcher representation about object. It is shown that the insufficient speed of hardware and the inaccessibility of some input parameters may lead to the big growth of errors at processing of experimental data.

Key words

Passive experiment, system-related problems, discrete-continuous model.

1 Introduction

The considerable attention in scientific physics community (European, Russian, American, etc. magazines) always has been given to methodological issues of physical experiment.

Features of active and passive experiment can be determined in accordance with Table 1.

Active experiment is carried out with application of artificial influence on object under the special program. It allows to solve research problems faster and more efficiently, but is more complex, requires great material expenses and can prevent the normal course of technological process.

At **passive experiment** the object functions in a usual mode. The information on object is registered in the form of signals from input and output variables. These signals can only be observed, but it is impossible to influence them, so the experimenter is in position of the passive observer [Bendat, Piersol,1966].

We will dwell on some specifics of passive experiment and on subtleties of data processing with aid of available mathematical methods.

Table 1. Experimental data processing system

Active experiment	**Passive experiment**
Creating a "refined conditions"	Virtually unrestricted dataflow
Minimizing the number of experiments	Narrow range of measurements
Maximum simplification of procedures of data processing	Strong noise
	Work in real time

2 System restrictions of the passive experiment

It is advised to put some text between a section heading and a subsection heading.

The use of data of normal operation of the object causes the wide range of system-related problems:

1. Work in real time.
2. Stochasticity of influences.
3. Dynamic properties of objects.
4. Limited observation intervals.
5. Discrete-continuous transformations.

About influence of some of these problems on the identification task solution we reported in S. St-Petersburg [Serdyuk, Troyanovskyi, 2009]. Here let's consider how such problems become apparent when the physical experiment processing.

Let's start from calculating the average estimate of a sample of limited amount. It is widely known the result of classical probability theory about reduce of the variance in N times relative to dispersion of equivalent independent variables (the number of N).

Therewith the dispersion level decrease of average is more complicated and depends on the correlation of data and the sampling period from a private realization.

Confidence intervals (Fig. 1) may differ [Troyanovskyi, 2004] in the case of independent signal and correlated signal. Note that the central limit theorem

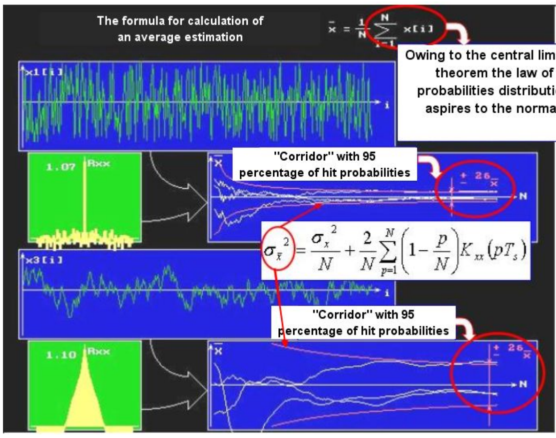

Figure 1. Fluctuations of the estimated average over realizations of limited length with different correlations

"works" by averaging a large number of samples, so that the boundaries of corridors are almost independent on the type of distribution of the original signal amplitudes.

Therewith averaging of the correlated signal increases a dispersion of a computed estimate, and speed of convergence of current average to expectation of signal become worse depending on degree of signal correlation.

It can lead to unexplained and unstable discrepancies of the properties of calculated estimates with the theoretically predicted.

3 The problem of using discrete-continuous model of a continuous process

A more complicated situation arises as result of the discrete-continuous transformations of signals with digital computer aid for continuous processes. Possibility and legitimacy of consideration of all processes (and continuous and discrete) in uniform time are considered in [Troyanovskyi, 2004]. However we would like to analyze especially: how the parity of the sampling period, time parameters of signal's properties and dynamic properties of object in common influence representation of the researcher about object?

Let the test object or process has a linear structure (Fig. 2).

Here

k – static transfer constant (gain factor);

$h(t)$ – object weight function, reflected its dynamic properties. The function is normalized

$$\int_{\lambda=0}^{\infty} f(\lambda)\,d\lambda \tag{1}$$

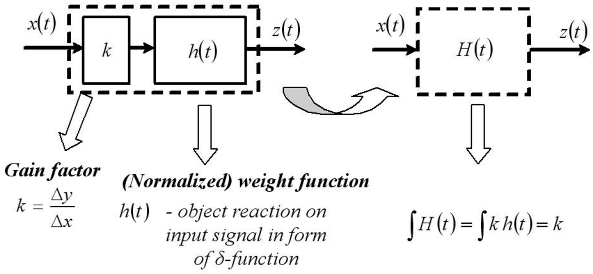

Figure 2. An object with linear structure

Connection of signals at the input and output of a linear dynamic object is described by the convolution equation

$$z(t) = k \int_{\tau=-\infty}^{t} x(\tau)h(t-\tau)\,d\tau \tag{2}$$

Remarks:

1. The last equation does not impose any restrictions on the type of signal, including the random signals.
2. Equation (2) is invariant to the time reference point.

It is easy to show that the static transfer constant k has the physical meaning of the ratio of steady output signal to the input step signal (at the end of the transition process), and the weight function $h(t)$ – an object reaction on input signal in form of δ-function at $k = 1$.

Note that if in the result of some experiment we can determine the generalized characteristic

$$H(t) = kh(t), \tag{3}$$

then by ratio (1)

$$\int H(t) = \int kh(t) = k. \tag{4}$$

Let us now see what happens when you use of a discrete-continuous model instead of a real continuous object. To do this, from the original continuous signal discrete samples are selected and subjected to digital processing and subsequent recovery (Fig. 3).

The process of sampling is described by the procedure of multiplying the original continuous signal on sequence δ- functions:

$$x^*(t) = x(t) \sum_{i=-\infty}^{\infty} \delta(t - iT_s), \tag{5}$$

where

$x^*(t)$ – quantified in time signal $x(t)$;

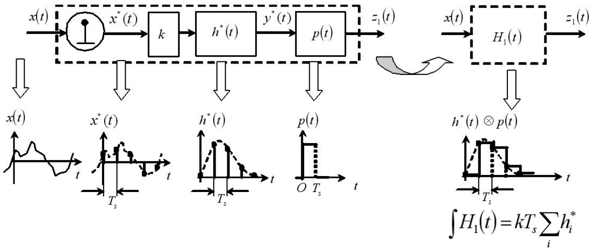

Figure 3. The structure of a discrete-continuous model

T_s – period of sampling.

For the linear case, the discrete output signal are determined with aid weighting factors of processing function and the sampled input signal as

$$y[i] = \sum_j h[j] \cdot x[i-j]. \tag{6}$$

If the discrete input and output signals are expressed as continuous functions of time, just as was done for the signal $x^*(t)$, the latter expression takes the form:

$$y^*(t) = \sum_j h[j]x(t) \sum_j \delta(t - iT_s - jT_s) \tag{7}$$

or

$$\begin{aligned} y^*(t) &= \int_\lambda h(\lambda) \sum_j \delta(\lambda - jT_s) \times \\ &\times x(t) \sum_j \delta(t - iT_s - \lambda) d\lambda = \\ &= \int_\lambda h^*(\lambda) x^*(t-\lambda) d\lambda, \end{aligned} \tag{8}$$

where it is defined

$$h^*(\lambda) = h(\lambda) \sum_j \delta(\lambda - jT_s) \tag{9}$$

discrete weight function of a linear processing unit, obtained by discretization of the original continuous function – a prototype, or the multiplication of discrete ordinates $h[j]$ on the sequence of δ-functions.

Restoring a continuous waveform from its discrete samples is made by a variety of ways. In the case of a simple digital-to-analog converter or the zero-order clamp, the last signal value, converted to analog form, remains at all near term T_s. It can be described by a weight function

$$p(t) = \begin{cases} 1, & at\, 0 \leq\ t <\ T_s \\ 0, & beyond\, this\, interval \end{cases} \tag{10}$$

The generalized weight function $H_1(t)$ and the reconstructed signal $z_1(t)$ are characterized by step functions (Fig. 3).

The difference between functions $H(t)$ and $H_1(t)$ determines the difference of static transfer constant of a real object, and discrete-continuous model.

Indeed, in the case of a continuous object

$$\int H(t) = \int kh(t) = k, \tag{11}$$

but for a model

$$k_1 = \int H_1(t) = kT_s \sum_i h[i] \tag{12}$$

that demonstrates a shift in the transfer constant k_1 relative to the true quantity k.

Simulation shows that the relative size of distortion is greater than 1 for aperiodic link of the first order and the relative size of distortion less than 1 for aperiodic link of the second and higher order. **The size of relative displacement can make tens and hundreds percent.** It depends on the weight function of the object and the ratio between period T_s and the weight function length of the object.

Thus, if the speed of digital computers begins to noticeably inferior temporal scales of the process, discrete-continuous model of the process causes significant distortion, even in a static transfer constant.

The same words can be said about the dynamic characteristics of the model, but the corresponding analysis, some results of which are described in [Troyanovskyi, 2004; Troyanovskyi, 2009], and the weight function length of the object.

4 Features of the experiment in the study of a multidimensional object

Let's consider the process of data processing, for example, when trying to determine the dynamic properties of the object.

As it is shown in [Serdyuk, Troyanovskyi, 2009; Troyanovskyi, 2004], for a one-parameter object with independent additive noise in the output signal an estimate of the transfer constant and the weight function of the object can be calculated, and their statistical properties can be determine. These results can be extended to the case of a multidimensional (multi-input) object with independent inputs, if there is a simultaneous consideration of all input signals. Indeed, in this case the generalized weight vector function of the object takes the form of a column, consisting of vector functions of individual channels and the total covariance matrix of input signals becomes strictly diagonal structure.

All this makes it possible to apply successfully the approach [Serdyuk, Troyanovskyi, 2009; Troyanovskyi, 2004], and dispersion of the output noise is equal to the dispersion of the initial noise

$$(\sigma^2_{out})_1 = \sigma^2_n. \quad (13)$$

Note that the efficiency of processing of the increased data flow can be enhanced through the organization of parallel computations [Serdyuk, Troyanovskyi, 2008].

However, an attempt to separate definition of the dynamic properties of individual channels leads to a deterioration of results. Here unrecorded channels play a role of additional noise (Fig. 4). They operate as additional unknown signals and increase an active noise.

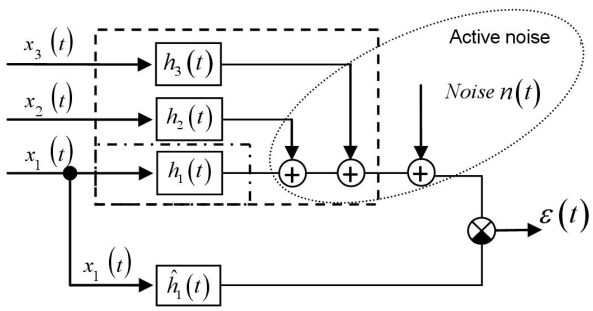

Figure 4. The scheme of separate identification for each individual channel

Indeed, for a 3-dimensional object level of current output noise is defined as

$$\left(\sigma^2_{out}\right)_1 = \sigma^2_n + \sigma^2_2 + \sigma^2_3. \quad (14)$$

If $\sigma^2_1 = \sigma^2_n = \sigma^2_2 = \sigma^2_3$, then

$$\begin{aligned}\left(\sigma^2_{out}\right)_1 &= \sigma^2_n + \sigma^2_2 + \sigma^2_n = \sigma^2_n + 2\sigma^2_0 = \\ &= \sigma^2_n \left(1 + \frac{2\sigma^2_0}{\sigma^2_n}\right) = \sigma^2_n \left(1 + \frac{2}{\gamma}\right). \end{aligned} \quad (15)$$

where

$$\gamma = \frac{\sigma^2_n}{\sigma^2_0} \quad (16)$$

– relative level of additive noise.

With independent input signals estimate of weigh function remains unbiased. However, growth in noise level leads to a deterioration of statistical stability of estimate, as the acting output noise for this case increases in $\left(1 + \frac{2}{\gamma}\right)$ times, i.e. tens and hundreds of percent, depending on the ratio of the dispersion of signals and noises.

In the more general case N-dimensional object ratio g of existing output noises in simultaneous and separate determination of the dynamic properties of individual channels is defined as

$$\begin{aligned} g = \frac{\left(\sigma^2_{out}\right)_2}{\left(\sigma^2_{out}\right)_1} &= \frac{\sigma^2_n + \sum\limits_{i=1}^{N-1} \sigma^2_i}{\sigma^2_n} = \\ &= \sigma^2_n \left(1 + \frac{\sum\limits_{i=1}^{N-1} \sigma^2_i}{\sigma^2_n}\right), \end{aligned} \quad (17)$$

and it leads to corresponding differences in dispersions of ordinate estimates of the weight function. Taking into account results of [Serdyuk, Troyanovskyi, 2009; Troyanovskyi, 2004] it is easy to show the following. At separate identification of channels there is an increase in a dispersion of estimations of weight functions. In case of channels equal in rights with identical dispersions of signals on an exit this increase reaches sizes

$$\tilde{g} = 1 + \frac{N-1}{\alpha\gamma} \quad (18)$$

Here the factor α reflects decrease in level of noise influence on identification accuracy in case of correlated noise).

Simulation results (Fig. 5) show how application of system approach and appropriate data processing algorithms can increase identification accuracy by tens and hundreds of percent as the view of the increasing number of channels.

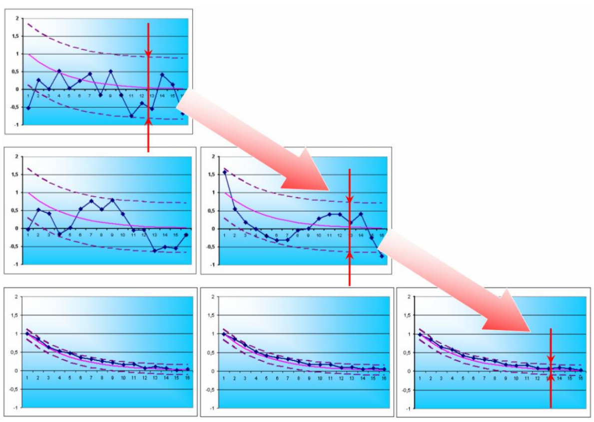

Figure 5. Growth of identification accuracy at the expense of parallel computations

Unfortunately, very often at the processing of physical experiments for multi-dimensional studied objects only a portion of input signals is known (or available for simultaneous measurements). Accordingly, in this case the experimenters have to pay for the duration of the experiment and increasing data amounts to achieve the desired statistical accuracy.

5 Conclusion

1. Passive physical experiment requires taking into account a wide range of system-related problems and some development of the methods of data analysis.
2. Involvement of the stochastic processes theory allows to avoid unexplained and unstable discrepancies of properties of the estimates of the average with the theoretical predictable on the basis of the classical theory of probability.
3. Analysis in time domain, involving the description of properties of a linear dynamic object on the basis of a convolution equation shows that such system restrictions as the speed of hardware involved and the inaccessibility of some input parameters may lead to the big growth of errors at processing of experimental data.

References

Bendat J. S. and Piersol A. G. (1966). *Random data: analysis and measurement procedures.* WileyInterscience. New York.

Serdyuk O.A., Troyanovskyi V.M. (2008). Identification problem and parallel computations. *Proceedings of the Fourth International Conference "Parallel Computations and Control Problems. PACO08"*. Moscow. October 28-29. SBN 978-5-91450-016-7. CD-ROM. pp. 1042–1048.

Serdyuk O.A., Troyanovskyi V.M. (2009). The ideal representations, rocks and realities of statistical methods' identification In *2009 IEEE International Conference on Control Applications.* - IEEE Catalog Number CFP09CCA-CDR, ISBN: 978-1-4244-4602-5, pp. 1472–1476.

Troyanovskyi V.M. (2004). *Informatsionno-upravlyayushchie sistemi i prikladnaya teoriya sluchaynikh protsessov.* Gelios ARV, Moscow.

Troyanovskyi V.M. (2009). Stationary and ergodicity analysis in discret/continuous stochastic systems as a methodical basis for the identification. *VIII International conference "System identification and control problems"*. Moscow, January 26–30. CD-ROM (1042–1048).