

Estimates Conformity Principle in the Problems of Identification

Vladimir Fursov

Image Processing Systems Institute, the Russian Academy of Sciences,
443001, Molodogvardeiskay, 151, Samara, Russia,
fursov@smr.ru

Abstract. A non-traditional approach to solving the identification problems using a small number of observations is developed. The approach is formulated as an estimate conformity principle. The problem of identifying a linear dynamic model is treated within this approach. A significant feature of the above method is the use of a quality criterion not requiring the knowledge of a priori probability noise distributions. The method is well suitable for real-time object identification in adaptive control systems.

1 Introduction

In control systems, the improvement of the system quality is, as a rule, achieved via the adaptation to the varying conditions of operation. The requirement that an adaptive system should have fast response to the changing environment makes us, if possible, use simple linear models and a small number of observations. Because of this, a number of estimation problems solved with the aim of the aircraft adaptation reduce to the following problem. One should find an estimate $\hat{\mathbf{c}}$ of the vector of parameter \mathbf{c} , using the $N \times M$ - matrix \mathbf{X} and the $N \times 1$ - vector \mathbf{y} ($N > M$) accessible from the direct observation or generated in one way or another, the matrix and the vector being related via the equation of linear regression

$$\mathbf{X}\mathbf{c} = \mathbf{y} + \boldsymbol{\xi}, \quad (1)$$

where $\boldsymbol{\xi}$ is the $N \times 1$ - vector of the “output” errors caused by the measurement errors and by the limitation imposed on the order of the model. In particular, a nonstationary linear discrete system with one input $u(k)$ and one output $y(k)$ is described by the equation [1]:

$$\mathbf{x}_{k+1} = \boldsymbol{\Phi}_k \mathbf{x}_k + \mathbf{d}_k u_k, \quad (2)$$

$$y_k = \mathbf{h}^T \mathbf{x}_k + v_k, \quad (3)$$

where \mathbf{x}_k is the n -dimensional vector of state, $\boldsymbol{\Phi}_k$ is the $n \times n$ - matrix of transient state, \mathbf{d}_k and \mathbf{h} are $n \times 1$ - vectors, and v_k is the random noise of measurements.

We assume that the time of formation of the observation records for identification is small, so that the parameters Φ_k and \mathbf{d}_k can be considered as being constant. In that case, for each row of the system (2) after N observations we can put down Eq. (1) in which the $M \times 1$ - vector of estimated parameters will include n elements of the corresponding row of the Φ matrix and one element of the \mathbf{d} vector, i.e. $M=n+1$. Solving Eq. (1) n times we can construct estimates for all elements of the Φ matrix and the \mathbf{d} vector.

From the discrete equations of state, Eqs. (2) and (3), (assuming, as before, that Φ and \mathbf{d} are constant) by a conventional way [1] we can go over to a single finite-difference equation:

$$y(k+1) = \sum_{i=1}^n a_i y(k-i+1) + \sum_{j=1}^n b_j u(k-j+1) + \xi(k+1), k = n-1, n-2, \dots \quad (4)$$

where $\xi(k) = v(k) + \sum_{i=1}^n a_i v(k-i)$ is the equivalent noise of output measurements. In this case, it is quite clear how to form the \mathbf{X} - matrix and the \mathbf{y} - vector in Eq. (1) from the input, $u(k)$, and output $y(k)$, measurements. The $M \times 1$ - vector of the sought parameters takes the form $\mathbf{c}^T = [a_1, \dots, a_n, b_1, \dots, b_n]$, i.e. at $b_1 \neq 0, M=2n$.

It is common practice to use the statistical theory of estimation for the development of identification algorithms [2]. With this approach, the quality criterion is defined on the basis of a priori information about the distribution of errors. If such an information is lacking a certain probable a priori hypothesis about error distribution is put forth. Then, the procedure constructed on its basis is statistically tested in order to concretize the probability model. Such an approach comes into conflict with the idea of system adaptation. In that case even “good” asymptotic properties of estimates in a particular situation may be misleading.

We develop an algebraic approach to the construction of estimation algorithms. This involves a more realistic formulation of the task as compared with that adopted in the theory of statistical estimation. We discuss the feasibility of solving such problems on the basis of the *estimates conformity principle*

2 Conformity Principle

The requirement to derive an estimate vector $\hat{\mathbf{c}}$ possibly close to the parameter vector \mathbf{c} implies that there exists the exact model of the system under estimation corresponding to Eq. (1)

$$\mathbf{y}^* = \mathbf{X}\mathbf{c}, \quad (5)$$

where

$$\mathbf{y}^* = \mathbf{y} - \boldsymbol{\xi}. \quad (6)$$

Obviously, Eq. (5) also holds for all sub-systems of lesser dimensionality constructed from the rows taken in an arbitrary combination. Of course, it appears impossible to attain the exact fulfillment of the equality in (5) for all the rows of the over-defined system because the error vector ξ in the initial set (1) is not known.

However, it is possible to specify a set $\{\tilde{\xi}_k\}$ of so-called correcting vectors $\tilde{\xi}_k$ with a view to selecting among them one sufficiently close to ξ . To the set of vectors $\{\tilde{\xi}_k\}$ there will correspond a set of variants of the corrected systems given by

$$\tilde{y}_k = \mathbf{X}\mathbf{c} + \xi - \tilde{\xi}_k, \quad k = \overline{1, K}, \tag{7}$$

where $\tilde{y}_k = \mathbf{y} - \tilde{\xi}_k$.

For each of the above variants, in its turn, one can specify L variants of the set of equations of lesser dimensionality $P: M < P < N$, with the $P \times N$ -matrix \mathbf{X}_l and $P \times 1$ - vectors $\tilde{y}_{k,l}$ and $\xi_l - \tilde{\xi}_{k,l}$:

$$\tilde{y}_{k,l} = \mathbf{X}_l\mathbf{c} + \xi_l - \tilde{\xi}_{k,l}, \quad k = \overline{1, K}, \quad l = \overline{1, L}. \tag{8}$$

For each (k,l) -variant one can also construct a set of sub-systems of lesser dimensionality, for example, with M rows of the set of equations (8) taken in an arbitrary combination:

$$\tilde{y}_{k,l,q} = \mathbf{X}_{l,q}\mathbf{c} + \xi_{l,q} - \tilde{\xi}_{k,l,q}, \quad q = \overline{1, Q}. \tag{9}$$

Here, index q serves to denote the sub-system number and the corresponding-sub-matrix and vectors of the k,l -th variant. If in constructing all sub-systems one takes exactly M rows, then $Q = C_p^M$.

Suppose also that

$$\text{Rank } \mathbf{X}_{l,q} = M, \quad q = \overline{1, Q} \tag{10}$$

for any matrix $\mathbf{X}_{l,q}$. Otherwise, it is always possible to ensure the fulfillment of this requirement by excluding the linearly dependent rows.

The essence of the estimate conformity principle is as follows. For each q -th sub-system of the $q = \overline{1, Q}$ k,l -th variant the estimate vector $\hat{\mathbf{c}}_{k,l,q}$ is calculated. We specify the function $W_l[\hat{\mathbf{c}}_{l,q}]$ characterizing the mutual closeness of the solutions $\hat{\mathbf{c}}_{k,l,q}$ derived on the k,l -th variant and the selection criterion of the most appropriate variant using the calculated closeness functions $W_{k,l}[\hat{\mathbf{c}}_{k,l,q}]$, $k = \overline{1, K}$, $l = \overline{1, L}$. The selection criterion of the point estimate of the sought-for parameter on a set of estimates corresponding to the selected variants and eventually used for constructing the sought-for point estimate $\hat{\mathbf{c}}$ is also specified.

Within the conformity principle, we can specify various functions of mutual closeness, the criteria for selecting the "best" variants based on the mutual closeness functions, and the criteria for constructing the estimate using the selected variants. We

consider the situation when the function of mutual closeness of the set of estimates on the variant is given by

$$W_i[\hat{\mathbf{c}}_{l,q}] = \sum_{i=1}^M \sum_{q=1}^Q (\hat{c}_{q,i} - \bar{c})^2, \tag{11}$$

where $\bar{c}_i = \frac{1}{Q} \sum_{q=1}^Q \hat{c}_{q,i}$

is the i -th component of the vector $\bar{\mathbf{c}}$ calculated by averaging the estimates $\{\hat{\mathbf{c}}_{l,q}\}$ derived on the variants for which the values of the mutual closeness function $W_i[\hat{\mathbf{c}}_{l,q}]$ have turned out to be less than a preset threshold.

Since the function of mutual closeness of estimates, $W[\hat{\mathbf{c}}_q]$, is not directly related to measuring and estimating errors a natural question arises: whether the application of the estimates conformity principle may guarantee a required accuracy when solving the set (1). To get the answer the qualitative analysis of possible situations has been conducted.

If $\tilde{\xi} = \mathbf{X}\mathbf{a}$, where \mathbf{a} is an arbitrary $M \times 1$ -vector, which implies that the correcting vector $\tilde{\xi}$ is a linear combination of columns of the matrix \mathbf{X} , then the component of the estimating error belonging to the null-space $N(\mathbf{X}^T)$ of the matrix \mathbf{X}^T [3] does not depend on the correcting vector and the estimating error will be affected only by the initial data. This means that in this situation the criterion (8) will not respond to variations in the corrections $\tilde{\xi}$.

In the general case, when the correcting vector $\tilde{\xi}$ also contains the component belonging to $N(\mathbf{X}^T)$ the estimates for different sub-systems will be different. This will result in non-zero values of the criterion (11). For a sequence of the correcting vectors, $\{\tilde{\xi}_k\}$, from $\tilde{\xi}_k \rightarrow \xi$ it follows that $Q(\tilde{\xi}_k) \rightarrow 0$. However, the fact that the criterion (11) converges to zero does not guarantee that the sequence of the correcting vectors will converge to the error vector ξ .

In view of the aforementioned properties we consider one possible strategy for constructing the estimates using the criterion (11). In particular, we discuss a method with which the estimates are being constructed for several variants of less-dimension sub-systems at $\tilde{\xi} = 0$. For each of them, the criterion (11) is calculated with the use of still less-dimension sub-systems. From the resulting set of sub-system variants and the corresponding estimates we select that for which the criterion value turns out to be the least one. In this paper, we give an example of implementation of the above-described procedure that demonstrates that the Least Square Method's (LSM) estimates can be essentially improved.

3 Analysis of Feasibility of Constructing the Estimates Using Non-statistical Criteria

Since the function of mutual closeness of estimates, $W[\hat{\mathbf{c}}_q]$, is not directly related with measurement and estimation errors, a natural question arises if the use of the estimate conformity principle guarantees that the set (1) is solved to the required accuracy. To get the answer we shall conduct a qualitative analysis of possible situations.

Consider the set (8). If $\text{Rank}(\mathbf{X})=M$, the vector $\xi - \check{\xi}$ appearing in the right-hand side of Eq. (8) can be represented as two components:

$$\xi - \check{\xi} = \mathbf{X}\Delta\mathbf{c} + \mathbf{T}_0\Delta\mathbf{d} , \tag{12}$$

где $\Delta\mathbf{c} = \mathbf{F}\Lambda^{-1/2}\mathbf{T}_\lambda^T(\xi - \check{\xi}) , \tag{13}$

$$\Delta\mathbf{d} = \mathbf{T}_0^T(\xi - \check{\xi}) . \tag{14}$$

Here \mathbf{T} is the $N \times N$ matrix and \mathbf{F} is the $M \times M$ matrix such that

$$\begin{aligned} \mathbf{T}^T \mathbf{X} \mathbf{F} &= \mathbf{S}, \quad \mathbf{X} = \mathbf{T} \mathbf{S} \mathbf{F}^T, \\ \mathbf{X} \mathbf{X}^T &= \mathbf{T} \mathbf{S} \mathbf{S}^T \mathbf{T}^T = \mathbf{T} \begin{bmatrix} \Lambda & 0 \\ 0 & 0 \end{bmatrix} \mathbf{T}^T, \end{aligned} \tag{15}$$

where \mathbf{S} is a diagonal $N \times M$ -matrix composed of so-called singular numbers, s_i , $i=1, M$, and $\Lambda = \text{diag}(\lambda_1, \lambda_1, \dots, \lambda_M)$ is a diagonal matrix of eigen-values λ_i , being the corresponding singular numbers s_i squared.

In view of the decomposition in Eqs. (12)-(14), Eq. (8) can be rewritten as follows

$$\check{\mathbf{y}} = \mathbf{X}(\mathbf{c} + \Delta\mathbf{c}) + \mathbf{T}_0\Delta\mathbf{d} . \tag{16}$$

Premultiply both sides of Eq. (29) by \mathbf{X}^T . In view of the property $\mathbf{X}^T \mathbf{T}_0 = 0$, we have

$$\mathbf{X}^T \check{\mathbf{y}} = \mathbf{X}^T \mathbf{X}(\mathbf{c} + \Delta\mathbf{c}) \tag{17}$$

Whence it follows that the root-mean-square estimate calculated on all data,

$$\check{\mathbf{c}} = [\mathbf{X}^T \mathbf{X}]^{-1} \mathbf{X}^T \check{\mathbf{y}} = \mathbf{c} + \Delta\mathbf{c} , \tag{18}$$

contains an error (13) that depends only on the components of the difference vector $\xi - \check{\xi}$, belonging to $R(\mathbf{X})$.

Let us now analyze the solutions to the sub-systems in (9) composed from (8) using corrected data. In view of the decomposition in (12), each of the sub-systems can be represented as

$$\check{\mathbf{y}}_q = \mathbf{X}_q(\mathbf{c} + \Delta\mathbf{c}) + \mathbf{T}_{0,q}\Delta\mathbf{d}, \quad q = \overline{1, \Sigma} \tag{19}$$

where $\mathbf{T}_{0,q}$ is a sub-matrix of the \mathbf{T}_0 -matrix corresponding to the q -th sub-system, and $\Delta\mathbf{c}$, $\Delta\mathbf{d}$ are the same as in (13) and (14). By multiplying from the left both sides of (19) by \mathbf{X}_q^T , it can easily be found that as distinct from (18), the LSM estimate calculated on the portion of data (for the q -th sub-system) will contain not only the error vector $\Delta\mathbf{c}$ but also an extra error

$$\Delta\tilde{\mathbf{c}}_q = [\mathbf{X}_q^T \mathbf{X}_q]^{-1} \mathbf{X}_q^T \mathbf{T}_{0,q} \Delta\mathbf{d} = [\mathbf{X}_q^T \mathbf{X}_q]^{-1} \mathbf{X}_q^T \mathbf{T}_{0,q} \mathbf{T}_0^T (\boldsymbol{\xi} - \tilde{\boldsymbol{\xi}}). \tag{20}$$

This is because for sub-systems of dimensionality lesser than N the product $\mathbf{X}_q^T \mathbf{T}_{0,q}$ is not necessarily equal to zero. Based on the aforementioned relationships, we can formulate the following important properties of the criterion in (11).

If $\tilde{\boldsymbol{\xi}} = \boldsymbol{\xi}$, in accordance with (13) and (14), $\Delta\mathbf{c}$ and $\Delta\mathbf{d}$ become zero simultaneously, and as one would expect, the sought-for vector parameter \mathbf{c} will be determined exactly both from the entire set of data and from data sets of all less-dimension sub-systems. The criterion in (11) that corresponds to the above-described situation takes its minimum value of zero.

If $\tilde{\boldsymbol{\xi}} = \mathbf{X}\mathbf{a}$ (where \mathbf{a} is an arbitrary $M \times 1$ -vector, that is the correcting vector $\tilde{\boldsymbol{\xi}}$ is a linear combination of the columns of matrix \mathbf{X}), then from the property $\mathbf{X}^T \mathbf{T}_0 = 0$ and Eq. (14) we have $\Delta\mathbf{d} = \mathbf{T}_0^T \tilde{\boldsymbol{\xi}}$, i.e. $\Delta\mathbf{d}$ does not depend on the correcting vector and the error in (20) will only be determined by the initial data. In this situation the criterion in (11) will not respond to variations in the corrections $\tilde{\boldsymbol{\xi}}$.

In the general case, when the correcting vector $\tilde{\boldsymbol{\xi}}$ comprises both the components belonging to the space of columns and the null-space, the error $\Delta\mathbf{c}$ appearing in (18) will be determined by the contribution of the component $\mathbf{X}\Delta\mathbf{c}$ in the right-hand side of (12). Whereas the sub-vectors $\mathbf{T}_{0,q}\Delta\mathbf{d}$ composed of the null-space component $\mathbf{T}_0\Delta\mathbf{d}$, will determine the set of errors in (20), which cause the estimation difference for different sub-systems. As a result, the value of the criterion in (11) will be non-zero. It is clear that for a sequence of correcting vectors $\{\tilde{\boldsymbol{\xi}}_k\}$ from $\tilde{\boldsymbol{\xi}}_k \rightarrow \boldsymbol{\xi}$ it follows that $Q(\tilde{\boldsymbol{\xi}}_k) \rightarrow 0$. However, the convergence of the criterion (11) to zero does not guarantee that the correcting vector will converge to the error vector.

Based on the above-formulated properties a number of strategies for constructing the estimates using the criterion in (11) have been proposed. In particular, in one method the estimates were constructed for several variants of less-dimension sub-systems at $\tilde{\boldsymbol{\xi}} = 0$. For each such a sub-system, the criterion (11) is calculated using still less-dimension sub-systems. From the resulting set of sub-system variants with the corresponding estimates the estimate characterized by the least criterion value is chosen.

4 Example

Below we discuss an example ($N=10, M=4$) of finding the solution using the criterion (11). The initial set of data was generated using a random number generator. Best conforming estimates were sought using $C_{10}^7=120$ sub-system variants, with 7 observations made for every variant. Then, for every variant, for $C_7^4=35$ (4×4)-sub-systems a set of LSM-estimates was calculated, for which the value of the criterion (11) was derived. The resulting set of the conformity criterion values was arranged in increasing order. For comparison, the criteria of estimation quality, $\|\Delta c\|$, were arranged in the same order.

Figure 1 shows the values of the conformity criterion. Figure 2 shows the corresponding values of the LMS-estimate quality criterion. The horizontal line shows the quality criterion value for the LSM-estimate calculated on all parameters.

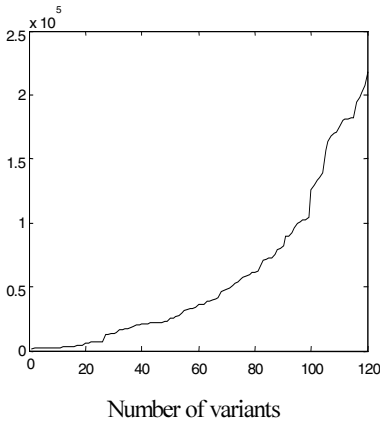


Fig. 1.

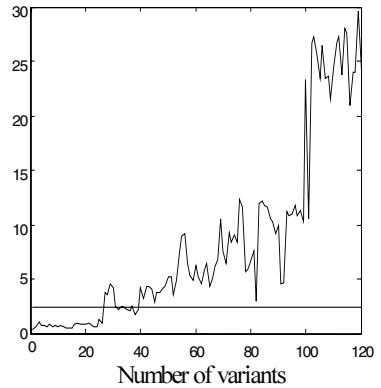


Fig. 2.

5 Conclusions

The aforesaid strategy for seeking the solution is conceptually similar to the ideas by P. Kalman discussed in Ref. [4]. In this case, the criterion (8) ensures the possibility of finding a sub-system with least-noised observations. The suitability of the strategy depends on how well the assumptions of existence of a noise-free sub-system are justified, as well as on the proper choice of dimension of the sub-systems that are assumed to include the above observations.

In order to find a least-noised sub-system it may require to search through a great number of variants. This fact is not astonishing, being the unavoidable fee for a priori information deficiency [5]. It seems reasonable that the application of simple (and, hence, relatively cheap) statistical schemes of data processing becomes feasible as a result of considerable expenses incurred while making a great number of observations at the stage of constructing the system (1).

Acknowledgement. This work was financially supported by the RF Ministry of Education, Samara Region Administration, the American Civilian Research and Development Foundation (CRDF Project SA-014-02) as part of the joint Russian-American program "Basic Research and Higher Education" (BRHE) and by the RFBR (grants N. 00-01-05001, N. 01-01-00097).

References

1. Graupe, D.: Identification of systems. Colorado State University Fort Collins, Robert E. Kriger Publishing Company Huntington, New York, 1978.
2. Ljung, L.: System Identification, Theory for the User. University of Linkoping, Sweden Prentice - Hall, Inc., 1987.
3. Bjorck Ake.: Least Squares Methods, Elsevier Science Publishers B.V. North Holland, 1990.
4. Kalman P.E.: Noised systems identification, Advances of Mathematical Sciences, v. 40, issue 4(244), 1985.
5. Fursov V.: Conformity principle in the problems of evaluating using a small number of observations, Proceedings of the IASTED International Conference Automation, Control, and Information Technology, June 10-13, 2002, Novosibirsk, Russia, p. 279–281.