S©MR ISSN 1813-3304

СИБИРСКИЕ ЭЛЕКТРОННЫЕ
МАТЕМАТИЧЕСКИЕ ИЗВЕСТИЯ

Siberian Electronic Mathematical Reports
http://semr.math.nsc.ru

Том 18, №2, стр. 792–804 (2021) УДК 519.6
DOI 10.33048/semi.2021.18.058 MSC 65F15, 41A30

OPERATOR-ORTHOREGRESSIVE METHODS FOR IDENTIFYING COEFFICIENTS OF LINEAR DIFFERENCE EQUATIONS

A.A. LOMOV

Abstract. We propose a new family of operator-orthoregressive methods for identifying the coefficients of linear difference equations from measurements of noisy solution at short time intervals. This family includes special cases of orthogonal regression (TLS) and variational identification (STLS) methods. The conditions of identifiability, as well as quantitative indicators of local identifiability, based on the numerical characteristics of the ellipsoids of deviations of the identified coefficients at small disturbances in measurements, are obtained. Computational algorithms are mentioned.

Keywords: linear difference equations, parameter identification, algebraic Fliess method, operator-orthoregressive method, orthogonal regression method, variational identification method, quantitative local identifiability indicators, Prony problem.

1. Introduction

Many problems of diagnostics, classification and forecasting are solved using ordinary linear differential equations with constant coefficients. Since stationarity is assumed over a finite time interval, methods with latent variables are used to solve related identification problems, in particular, the orthogonal regression (Total Least Squares) [1, 2, 3] and variational identification (Structured Total Least Squares) methods [4, 5]. Latent variables are introduced into objective functions to eliminate the influence of additive errors in variables (EIV).

Lomov, A.A., Operator-orthoregressive methods for identifying coefficients of linear difference equations.

The work is supported by RFFI (grant 19-01-00754).
Received December, 28, 2020, published July, 14, 2021.

In the last two decades, an alternative approach to EIV identification problems in a finite time has been actively developed. These are operator-algebraic methods based on integral transformations of measurements in order to obtain simple algebraic equations for the parameters to be identified. The idea of the integral transformations was used by many authors, see, for example, [6, 7, 8]. Among them, M. Fliess [9, 10] was apparently the first to propose transformations that are exactly reversible over a finite observation interval.

In [11], the operator-orthoregressive identification method was presented, which combines the EIV approach and integral transformations of the M. Fliess type. This combination made it possible to guarantee the consistency of the parameter estimates under natural excitation conditions without restrictions on stability and controllability of the model difference equation. Developing the results of [11], here we describe the set of all possible variants of the operator-orthoregressive method and show that it includes the TLS and STLS methods. Also, the conditions for global identifiability and quantitative indicators of local identifiability are obtained.

2. The algebraic method for difference equations

The history of operator identification methods can be counted from the algebraic method of M. Fliess [9]. For example, consider a linear autonomous homogeneous differential equation with real coefficients (similarly, we can study the more general case of systems of inhomogeneous equations):

$$x^{(n)} + a_{n-1}x^{(n-1)} + \ldots + a_0 x = 0, \tag{1}$$

$$x^{(k)} = x^{(k)}(t) \doteq \frac{d^k}{dt^k}x(t), \quad t \in [0, T).$$

Suppose a function $x(t)$ is given that satisfies equation (1) with unknown coefficients $a_0, \ldots, a_{n-1}$ to be identified. The problem becomes meaningful if the functions $x(t)$ are known inaccurately, i. e. when we have disturbed measurements $\check{x}(t) = x(t) + \Delta x(t)$ instead of exact solutions $x(t)$. M. Fliess [9, 10] used certain algebraic transformations to eliminate the influence of the initial conditions of an unknown exact solution $x(t)$ to obtain equations for the coefficients $a_0, \ldots, a_{n-1}$. Measurement derivatives are not used in the resulting equations, which increases the resistance of the method to random noise in measurements. In the method of M. Fliess, the equation (1) is replaced by a system of equivalent integral equations including the function $x(t)$ and some auxiliary functions $\mu_{ij}(t)$, which we call the Fliess polynomials [11]:

$$\begin{bmatrix} \langle \mu_{10}, x \rangle & \langle \mu_{11}, x \rangle & \ldots & \langle \mu_{1n}, x \rangle \\ \vdots & \vdots & & \vdots \\ \langle \mu_{n0}, x \rangle & \langle \mu_{n1}, x \rangle & \ldots & \langle \mu_{nn}, x \rangle \end{bmatrix} \begin{bmatrix} a_0 \\ \vdots \\ a_{n-1} \\ 1 \end{bmatrix} = 0, \tag{2}$$

$$\langle \mu_{ij}, x \rangle \doteq T^{n-j} \int_0^1 \mu_{ij}(\tau)\, x(T\tau)\, d\tau, \quad i \in \overline{1, n}, \quad j \in \overline{0, n}.$$

In practice, measurements $\check{x}$ are plugged directly into the equation (2) in the hope that integral convolutions will reduce the effect of noise. The statistically correct accounting for the measurement of the noise in the M. Fliess method was discussed in [11].

Now we pass from (1) to the difference equation, specifying the uniform time grid

$$\tau_1 = 0, \tau_2 = h, \dots, \tau_N = 1, \quad h \doteq \tfrac{1}{N-1},$$

and denoting $x_k \doteq x(T\tau_k)$. The time sequence $x \doteq \begin{bmatrix} x_1 & \dots & x_N \end{bmatrix}^{\mathrm{T}}$ is a solution to the difference equation with some constant coefficients α_i:

$$\text{(3)} \qquad \alpha_n x_{k+n} + \alpha_{n-1} x_{k+n-1} + \ldots + \alpha_0 x_k = 0, \quad k \in \overline{1, N-n}, \quad \alpha_n \neq 0.$$

We can write (3) in a matrix form:

$$\text{(4)} \qquad G_\alpha^{\mathrm{T}} x = 0, \quad \alpha^{\mathrm{T}} \doteq \begin{bmatrix} \alpha_0 & \alpha_1 & \dots & \alpha_n \end{bmatrix},$$

$$G_\alpha^{\mathrm{T}} \doteq \begin{bmatrix} \alpha_0 & \alpha_1 & \dots & \alpha_n & & & 0 \\ & \alpha_0 & \alpha_1 & \dots & \alpha_n & & \\ & & \ddots & \ddots & & \ddots & \\ 0 & & & \alpha_0 & \alpha_1 & \dots & \alpha_n \end{bmatrix} \doteq \backslash \alpha^{\mathrm{T}} \backslash_{(N-n)\times N}.$$

We define the shift matrices

$$\text{(5)} \qquad E_j^{\mathrm{T}} \doteq \begin{bmatrix} 0_j & I_{N-n} & 0_{n-j} \end{bmatrix} \in \mathbb{R}^{(N-n)\times N}, \quad j \in \overline{1, N-n},$$

where $0_j \in \mathbb{R}^{(N-n)\times j}$ are submatrices of zeros, and I_{N-n} is the identity matrix. Let

$$\text{(6)} \quad V \doteq \begin{bmatrix} x_1 & \dots & x_n & x_{n+1} \\ x_2 & \dots & x_{n+1} & x_{n+2} \\ \vdots & & \vdots & \vdots \\ x_{N-n} & \dots & x_{N-1} & x_N \end{bmatrix} = \begin{bmatrix} E_0^{\mathrm{T}} x & \dots & E_n^{\mathrm{T}} x \end{bmatrix} \doteq \begin{bmatrix} x_{[0]} & \dots & x_{[n]} \end{bmatrix}.$$

Then the following equalities are true:

$$\text{(7)} \qquad G_\alpha^{\mathrm{T}} = \alpha_0 E_0^{\mathrm{T}} + \ldots + \alpha_n E_n^{\mathrm{T}},$$

$$\text{(8)} \qquad G_\alpha^{\mathrm{T}} x = \left(\alpha_0 E_0^{\mathrm{T}} + \ldots + \alpha_n E_n^{\mathrm{T}}\right) x = V\alpha.$$

Consider the problem of obtaining analogues of integral equations (2) in the case of discrete time. We have to replace $\langle \mu_{ij}, x \rangle$ with products $\mu_{ij}^{\mathrm{T}} x$, where

$$\mu_{ij} \doteq \begin{bmatrix} \mu_{ij}[1] & \dots & \mu_{ij}[N] \end{bmatrix}^{\mathrm{T}} \in \mathbb{R}^N, \quad \mu_{ij}[k] \doteq \mu_{ij}(\tau_k)$$

(the same notations x, μ_{ij} are used both for the functions $x(t), \mu_{ij}(\tau)$ and for the corresponding time sequences composed of the values of the functions at grid points). For example, the trapezoid formula leads to the products

$$\langle \mu_{ij}, x \rangle \approx T^{n-j} h \tilde{\mu}_{ij}^{\mathrm{T}} x, \quad \tilde{\mu}_{ij}^{\mathrm{T}} \doteq \mu_{ij}^{\mathrm{T}} \operatorname{diag}\left(0.5, 1, \ldots, 1, 0.5\right),$$

and to approximate equations $\begin{bmatrix} \mu_{i0}^{\mathrm{T}} x & \mu_{i1}^{\mathrm{T}} x & \dots & \mu_{in}^{\mathrm{T}} x \end{bmatrix} a \approx 0$ instead of (2) [11]. We are interested in exact equalities. The following theorem describes all possible functions $\varphi_{ij} \in \mathbb{R}^N$, substitution of which instead of μ_{ij} leads to exact equalities.

Theorem 1. *Let $x \in \mathbb{R}^N$ be a solution to the equation (3) with a vector of coefficients $\alpha = \begin{bmatrix} \alpha_0 & \dots & \alpha_n \end{bmatrix}^{\mathrm{T}}$. Then for any set of linearly independent vectors*

$\{p_1, \dots, p_m\} \subset \mathbb{R}^{N-n}$ *the following system of equations holds*

$$\begin{bmatrix} \varphi_{10}^{\mathrm{T}} x & \dots & \varphi_{1n}^{\mathrm{T}} x \\ \vdots & & \vdots \\ \varphi_{m0}^{\mathrm{T}} x & \dots & \varphi_{mn}^{\mathrm{T}} x \end{bmatrix} \alpha = 0, \quad m \leqslant N-n, \tag{9}$$

where

$$\varphi_{ij} \doteq E_j p_i \tag{10}$$

and E_j *is the shift matrix* (5). *Conversely, if the system of equations* (9) *is valid for some set of linear independent vectors* $\{p_1, \dots, p_m\} \subset \mathbb{R}^{N-n}$ *and* $m = N-n$, *then the function* x *is a solution to the difference equation* (3).

The system of equations (9) can be considered as an integral (convolutional) analogue of the difference equation (3), in which instead of a function x, convolutions (scalar products) of a function x with functions φ_{ij} are involved. In turn, for each $i \in \overline{1, m}$ the functions φ_{ij} are obtained by j-shift of the zeros-complemented vector $p_i \doteq \begin{bmatrix} p_{i1} & \dots & p_{i,N-n} \end{bmatrix}^{\mathrm{T}}$ so that the columns $\varphi_{i0}, \dots, \varphi_{in}$ form the Toeplitz matrix

$$\begin{bmatrix} \varphi_{i0} & \dots & \varphi_{in} \end{bmatrix} = \begin{bmatrix} p_{i1} & & 0 \\ \vdots & \ddots & \\ p_{i,N-n} & & p_{i1} \\ & \ddots & \vdots \\ 0 & & p_{i,N-n} \end{bmatrix} = \begin{bmatrix} E_0 p_i & \dots & E_n p_i \end{bmatrix} \doteq \backslash p_i \backslash_{N \times (n+1)}. \tag{11}$$

Each set of linearly independent vectors $P = \{p_1, \dots, p_m\}$ in $\mathbb{R}^{N-n}$ defines a set of functions $\{\varphi_{10}, \dots, \varphi_{mn}\}$ (9), therefore, changing P, we enumerate all possible variants of the algebraic method for the equation (3).

The proofs of the theorem and subsequent statements are given in the appendix.

We choose $m \geqslant n$ and write (9) in a matrix form

$$Y\alpha = 0, \tag{12}$$

$$Y \doteq \begin{bmatrix} \varphi_{10}^{\mathrm{T}} x & \dots & \varphi_{1n}^{\mathrm{T}} x \\ \vdots & & \vdots \\ \varphi_{m0}^{\mathrm{T}} x & \dots & \varphi_{mn}^{\mathrm{T}} x \end{bmatrix} \in \mathbb{R}^{m \times (n+1)}. \tag{13}$$

The system (12) may be overdetermined, so its solution α is found by pseudo-inversion of the n-column submatrix of Y or in another way (see below), taking into account the presence of disturbances in the measurements x and the proximity of the equality (12) in this case.

3. Identifiability conditions

Here we describe the conditions for x that guarantee the uniqueness of the solution $\alpha = \begin{bmatrix} \alpha_0 & \dots & \alpha_{n-1} & 1 \end{bmatrix}^{\mathrm{T}}$ of the system of equations (12), (13).

Definition 1. *Some property* $\mathcal{B}$ *is said to hold for "almost all" elements of the open set* $\mathcal{A} \subset \mathscr{V}$ *in the normed vector space* $\mathscr{V}$, *if the set* $\mathcal{B} \subset \mathscr{V}$ *of elements with this property is everywhere dense in* $\mathcal{A}$, *that is, for any element* $a \in \mathcal{A}$ *and any arbitrarily small* ε *there exists an element* $a_\varepsilon \in \mathcal{A}$, *such that* $\|a_\varepsilon - a\| < \varepsilon$ *and* $a_\varepsilon \in \mathcal{B}$.

Lemma 1. *The property* $\det A^{\mathrm{T}}B \neq 0$ *holds for almost all matrices* $A, B \in \mathbb{R}^{m\times n}$, $m \geqslant n$ *of full rank* n.

Definition 2. *A solution* x *of the equation* (3) *will be called* complete *if it does not satisfy any equation of the form* (3) *of a lower order* $\nu < n$.

The equation (3) is equivalent to $G_\alpha^{\mathrm{T}} x = V\alpha = 0$ (cf. (4), (8)). We select the submatrix $\overline{V}(x)$ formed by the first n columns of the matrix $V = V(x)$ (6):

$$V = \left[\begin{array}{c|c} & x_{n+1} \\ & x_{n+2} \\ \overline{V}(x) & \vdots \\ & x_N \end{array}\right].$$

The solution x is complete if and only if the equality $\operatorname{rank} V(x) = \operatorname{rank} \overline{V}(x) = n$ holds.

Theorem 2. *For almost all matrices* $P \doteq \begin{bmatrix} p_1 & \dots & p_m \end{bmatrix} \in \mathbb{R}^{(N-n)\times m}$, $m \leqslant N - n$, *of full rank* m, *the following statement is true: the vector of coefficients* $\alpha = \begin{bmatrix} \alpha_0 & \dots & \alpha_{n-1} & 1 \end{bmatrix}^{\mathrm{T}}$ *is uniquely calculated from the system of equations* (12) *with* $\varphi_{i0}, \dots, \varphi_{in}$, $i \in \overline{1, m}$ (9) *if and only if* $\operatorname{rank} \overline{V}^{\mathrm{T}} \overline{V} = n$ *and* $m \geqslant n$.

Corollary 1. *Let* x *be a complete solution to equation* (3). *Then, to uniquely calculate the vector of coefficients* α *from the system of equations* (12) *with* $\varphi_{i0}, \dots, \varphi_{in}$, $i \in \overline{1, m}$ (9) *for almost all sets of linearly independent vectors* $\{p_1, \dots, p_m\} \subset \mathbb{R}^{N-n}$ *it suffices to satisfy the inequality* $m \geqslant n$.

The uniqueness condition "for almost all" matricies P of full rank in Theorem 2 leaves open the question of what happens in a small neighborhood of some P', x' for which the uniqueness condition is not satisfied. The existence of such P', x' is confirmed by the following example. Let

$$P' = \begin{bmatrix} p_1 & \dots & p_{N-2n} \end{bmatrix} = \backslash\alpha\backslash_{(N-n)\times(N-2n)}. \tag{14}$$

The matrix P' has the full rank $N-2n$ and at the same time $P'^{\mathrm{T}}\overline{V} = 0$, $\operatorname{rank} P'^{\mathrm{T}}\overline{V} < n$. This means that for any x the vector α cannot be uniquely determined by the equation $P'^{\mathrm{T}}V\alpha = 0$ or the equivalent equation $Y\alpha = 0$.

By Theorem 2, there always exists P_ε, which is slightly different from P' (14) and guarantees, given complete measurements x, that the condition $\operatorname{rank} P_\varepsilon^{\mathrm{T}}\overline{V} = \operatorname{rank} P_\varepsilon^{\mathrm{T}} V = n$ is satisfied, sufficient for identifiability of α. Thus, from a practical point of view, identifiability with complete measurements x is "always". The question is quantitative; namely, for which sets p_i (and also m, x in (12)) the vector α is identified "better" and in what sense; in particular, what happens when P is chosen close to P' (14).

The answer can be given by means of quantitative indices of local identifiability of α. They can be constructed by presenting the identified values (*the estimates*) of α as the minimum points of some objective function and expanding the objective function in terms of small perturbations in the exact solution x (see [12] and section 5).

4. Objective function

To select the objective function, we need to fulfill the necessary conditions: the uniqueness and continuity of the minimum point with small changes in x and its consistency in the sense of the convergence of the minimum point to the true value with an increase in the number of random measurements x. Let the measurements be specified as

$$\check{x}_{(l)} = x_{(l)} + \Delta x_{(l)}, \quad l = \overline{1, L},$$

where $x_{(l)}$ are functions obeying the equation (3) with an unknown true $\alpha = \alpha_*$ and $\Delta x_{(l)}$ are disturbances. Then instead of (12) we will have the system of equations for calculating the estimate $\hat{\alpha}$ of α_*:

$$\left(\check{Y} - \widehat{\Delta Y}\right)\hat{\alpha} \doteq \widehat{Y}\hat{\alpha} = 0, \tag{15}$$

where the following constructs are used:

$$\check{Y} \doteq Y\left(\check{x}_{(1)}, \ldots, \check{x}_{(L)}\right), \quad \widehat{\Delta Y} \doteq Y\left(\widehat{\Delta x}_{(1)}, \ldots, \widehat{\Delta x}_{(L)}\right),$$

$$Y\left(\check{x}_{(1)}, \ldots, \check{x}_{(L)}\right) \doteq \frac{1}{\sqrt{L}}\begin{bmatrix} Y(\check{x}_{(1)}) \\ \vdots \\ Y(\check{x}_{(L)}) \end{bmatrix}. \tag{16}$$

The processes $\widehat{\Delta x}_{(l)}$ are estimates of the disturbances $\Delta x_{(l)}$ and must be identified together with the vector $\hat{\alpha}$. They are treated as latent variables.

The system of equations (15) admits an infinite set of $\left(\widehat{\Delta Y}, \hat{\alpha}\right)$ pairs. To verify this, we choose some set of solutions $\left\{\hat{x}_{(1)}, \ldots, \hat{x}_{(L)}\right\}$ of the equation (3) with an arbitrary $\hat{\alpha}$; then all pairs $\left(\widehat{\Delta Y}, \hat{\alpha}\right)$ with elements

$$\widehat{\Delta Y} \doteq Y\left(\check{x}_{(1)} - \hat{x}_{(1)}, \ldots, \check{x}_{(L)} - \hat{x}_{(L)}\right)$$

by Theorem 1 are solutions to the system (15).

The only solution to (15) is sought with the use of the minimum condition for some objective function, for example, $\|\widehat{\Delta Y}\|_R^2$ (this is an orthogonal regression problem [2, 3, 13]):

$$\hat{\alpha} = \arg\min_{\alpha} J(\alpha), \quad J(\alpha) \doteq \min_{\Delta Y:\, \left(\check{Y} - \Delta Y\right)\alpha = 0} \|\Delta Y\|_R^2. \tag{17}$$

In the particular case of the relation (13) between ΔY and $\Delta x_{(l)}$, the estimate $\hat{\alpha}$ (17) is called operator-orthoregressive [11]. In it, in contrast to the classical problem of orthogonal regression, the measurements $\check{x}_{(l)}$ enter into the matrix $\check{Y}$ not directly, but through the convolutions (13) with the functions φ_{ij}.

We will use the $\otimes$ symbol for the Kronecker product and $^+$ for the Moore — Penrose pseudoinverse ($AA^+A = A$, $A^+AA^+ = A^+$, $A^{++} = A$).

Proposition 1. *Let the matrix norm* $\|Y\|_R^2$ *be defined by the equality* $\|Y\|_R^2 \doteq \|y\|_R^2 \doteq y^{\mathrm{T}}Ry$, *where* $R \geqslant 0$ *is a nonnegative definite symmetric matrix, and* $y \doteq \operatorname{vect} Y$ *is a row alignment,* $Y \in \mathbb{R}^{Lm\times(n+1)}$, $y \in \mathbb{R}^{Lm(n+1)}$. *Then, given the additional non-limiting condition* $\Delta y \in \operatorname{im} R$ *and existence of the inverse matrix* C_α, *the following equivalent representation of the objective function* (17) *is true:*

$$J(\alpha) = \check{y}^{\mathrm{T}}\Gamma_\alpha C_\alpha \Gamma_\alpha^{\mathrm{T}}\check{y} = \alpha^{\mathrm{T}}\check{Y}^{\mathrm{T}}C_\alpha\check{Y}\alpha, \tag{18}$$

$$C_\alpha \doteq \left(\Gamma_\alpha^{\mathrm{T}}R^+\Gamma_\alpha\right)^{-1}, \quad \Gamma_\alpha \doteq I_{Lm} \otimes \alpha. \tag{19}$$

Weighting matrix R. If it is assumed that the disturbances ΔY, $\Delta x_{(l)}$ are random with known properties, then for the correct value of R we can ensure statistical consistency in the sense of the convergence of the minimum point $\hat{\alpha}$ of the objective function $J(\alpha)$ to the true value: $\hat{\alpha} \underset{L\to\infty}{\to} \alpha_*$ [15]. Let the vector Δx have zero mathematical expectation $\mathbf{M}\,\Delta x = 0$ and the covariance matrix $\operatorname{cov}\Delta x = \sigma^2 I_N$. For the vector $\Delta y \doteq \operatorname{vect}\Delta Y$, from (13) we have

$$\Delta y \doteq \begin{bmatrix}\varphi_{10}^{\mathrm T}\Delta x & \dots & \varphi_{1n}^{\mathrm T}\Delta x & \dots & \varphi_{m0}^{\mathrm T}\Delta x & \dots & \varphi_{mn}^{\mathrm T}\Delta x\end{bmatrix}^{\mathrm T} \doteq \Sigma^{\mathrm T}\Delta x \in \mathbb{R}^{m(n+1)},$$

$$(20)\quad \Sigma \doteq \big[\underbrace{\varphi_{10} \quad \dots \quad \varphi_{1n}}_{\Sigma_1} \cdots \underbrace{\varphi_{m0} \quad \dots \quad \varphi_{mn}}_{\Sigma_m}\big] \doteq \begin{bmatrix}\Sigma_1 & \dots & \Sigma_m\end{bmatrix} \in \mathbb{R}^{N\times m(n+1)},$$

$$\Sigma_i = \backslash p_i \backslash_{N\times(n+1)}\,.$$

Then $\mathbf{M}\,\Delta y = 0$ and $\operatorname{cov}\Delta y = \sigma^2\Sigma^{\mathrm T}\Sigma$.

Theorem 3. *Let equations* (12), (16) *and measurements* $\check{y}_{(l)} = y_{(l)} + \Delta y_{(l)} = \Sigma^{\mathrm T}\left(x_{(l)} + \Delta x_{(l)}\right) \in \mathbb{R}^{m(n+1)}$, $l = \overline{1,L}$, *be given, where the variables* $\Delta y_{(l)}$ *are random, so that for different* l *they are statistically independent and equally distributed,* $\mathbf{M}\,\Delta y_{(l)} = 0$, $\operatorname{cov}\Delta y_{(l)} = \sigma^2\Sigma^{\mathrm T}\Sigma$. *We also assume that the undisturbed values* $y_{(l)} = \Sigma^{\mathrm T}x_{(l)}$ *satisfy the condition of identifiability (Theorem 2 and Corollary 1), that is, the equation* $Y\alpha = 0$ *implies* $\alpha = \alpha_*$. *Then for*

$$(21)\qquad R = I_L \otimes (\Sigma^{\mathrm T}\Sigma)^+ \in \mathbb{R}^{Lm(n+1)\times Lm(n+1)}$$

the estimate $\hat{\alpha}$ (17), (18) *is consistent in the sense of convergence* $\hat{\alpha} \underset{L\to\infty}{\to} \alpha_*$.

This theorem follows directly from Theorems 1 and 2 in [15], taking into account that the estimate $\hat{\alpha}$ (17), (18) is projective (or variational, orthoregressive) in terms of [5, 15].

Proposition 2. *For matrices* Γ_α (19) *and* R (21), *the equalities* $\Sigma\,(I_m \otimes \alpha) = G_\alpha P$ *and* $\Gamma_\alpha^{\mathrm T} R^+ \Gamma_\alpha = I_L \otimes (P^{\mathrm T} G_\alpha^{\mathrm T} G_\alpha P)$ *are true.*

Lemma 2. *If the columns of the matrix* P *are linearly independent, then the columns of the matrix* $\backslash\alpha\backslash\,P \doteq G_\alpha P$ *are also linearly independent.*

Corollary 2. *For matrices* Γ_α (19) *and* R (21), *there is always an inverse*

$$\left(\Gamma_\alpha^{\mathrm T} R^+ \Gamma_\alpha\right)^{-1},$$

and the following equalities hold:

$$(22)\qquad J(\alpha) = \sum_{i=1}^{L} \alpha^{\mathrm T} Y(\check{x}_{(i)})^{\mathrm T} C_\alpha Y(\check{x}_{(i)})\alpha,$$

$$C_\alpha \doteq \left(P^{\mathrm T} G_\alpha^{\mathrm T} G_\alpha P\right)^{-1}.$$

The objective functions (18), (22) are closely related [5] with the objective function of the modified Prony problem [17, formula (7)], [18]. Computational algorithms for minimizing these objective functions were discussed in [11, 16].

5. Quantitative indicators of local identifiability

To simplify the formulas, let us restrict ourselves to the case $L = 1$. We will use the affine parameterization

$$\alpha = D\theta + d = \underbrace{\begin{bmatrix} D & d \end{bmatrix}}_{\mathscr{D}} \underbrace{\begin{bmatrix} \theta \\ 1 \end{bmatrix}}_{\vartheta} \doteq \mathscr{D}\vartheta, \tag{23}$$

where $\theta \in \mathbb{R}^v$, $v \leqslant n$, is the vector parameter on which the coefficients α_i in (3) depend, the matrix $D \in \mathbb{R}^{(n+1)\times v}$ and the column $d \in \mathbb{R}^{n+1}$ are chosen such that the columns of the composite matrix $\mathscr{D} \doteq \begin{bmatrix} D & d \end{bmatrix}$ are linearly independent. If $v = n$ and $\mathscr{D} = \begin{bmatrix} I_n & 0 \\ 0 & \alpha_n \end{bmatrix}$, then we get the simplest parameterization $\theta = \begin{bmatrix} \alpha_0 & \ldots & \alpha_{n-1} \end{bmatrix}^{\mathrm{T}}$ with a fixed value of α_n. To make the order of the equation (3) independent of θ, it is assumed that on the given compact set $\theta \in \Theta \subset \mathbb{R}^v$, the leading coefficient α_n is nonzero.

Taking into account (21), (22), and (23), the objective function $J(\alpha)$ (18) can be expressed as

$$J(\alpha) = \alpha^{\mathrm{T}} \check{Y}^{\mathrm{T}} C_\alpha \check{Y} \alpha = \vartheta^{\mathrm{T}} \mathscr{D}^{\mathrm{T}} \check{Y}^{\mathrm{T}} C_\theta \check{Y} \mathscr{D} \vartheta, \tag{24}$$

$$C_\theta \doteq \left(P^{\mathrm{T}} G^{\mathrm{T}}_{\alpha(\theta)} G_{\alpha(\theta)} P \right)^{-1}.$$

We define the *sensitivity matrix*

$$S^{\mathrm{T}} \doteq -\left(J''_{\theta\theta}\right)^{-1} J''_{\theta y}$$

using partial derivatives of the objective function (18)

$$J(\theta, y) \doteq J(\alpha(\theta), y) = \alpha^{\mathrm{T}} Y^{\mathrm{T}} C_\alpha Y \alpha = y^{\mathrm{T}} \Gamma_\alpha C_\alpha \Gamma^{\mathrm{T}}_\alpha y, \quad \alpha'_\theta = D$$

with respect to θ, y at the «true» point $J(\theta, y) = 0$. The following statement clarifies the meaning of the matrix S, see also [12, Remark 1].

Claim. In the limiting case $\sigma \to 0$, the ball $B_\sigma \doteq \{\|\Delta x\| \leqslant \sigma\} \subset \mathbb{R}^N$, to which disturbances belong, corresponds to the ellipsoid of deviations of the variables $y = S^{\mathrm{T}} x$

$$\{\Delta y = \Sigma^{\mathrm{T}} \Delta x : \ \|\Delta x\| \leqslant \sigma\} = \left\{\Delta y : \ \Delta y^{\mathrm{T}} \left(\Sigma^{\mathrm{T}} \Sigma\right)^{-1} \Delta y \leqslant \sigma^2 \right\} \leftrightarrow B_\sigma$$

and the deviation ellipsoid

$$E_\sigma = \left\{\Delta\theta : \ \Delta\theta^{\mathrm{T}} \left(S^{\mathrm{T}} \Sigma^{\mathrm{T}} \Sigma S\right)^{-1} \Delta\theta \leqslant \sigma^2 \right\} \leftrightarrow B_\sigma$$

of the parameter estimates

$$\hat{\theta} = \arg\min_\theta J(\theta, y + \Delta y) = \theta + \Delta\theta. \tag{25}$$

To compare the ellipsoids of deviations of estimates obtained by different methods, one of the numerical characteristics of the *deviation matrix* $\mathscr{R} \doteq S^{\mathrm{T}} \Sigma^{\mathrm{T}} \Sigma S$ can be used. For example, the largest eigenvalue $\lambda_{\max}(\mathscr{R})$ (square of the major semiaxis length of the E_σ), determinant $\det \mathscr{R}$ (volume E_σ), etc. [19].

Proposition 3. *For deviations of estimates* $\hat{\theta}$ (24), (25) *at the "true" point* $J(\theta, y) = 0$*, the deviation matrix is*

$$\mathscr{R} = \left(D^{\mathrm{T}} Y^{\mathrm{T}} C_\theta Y D\right)^{-1}. \tag{26}$$

Note that the expression for the deviation matrix $\mathscr{R}$ (26) contains the Y and C_θ matrices, which depend on the choice of $P = \begin{bmatrix} p_1 & \dots & p_m \end{bmatrix}$ (see (10), (13), (24)). In the complete nondegenerate case $P = \begin{bmatrix} p_1 & \dots & p_{N-n} \end{bmatrix}$, $\det P \neq 0$, this dependence disappears. This fact follows from the invariance of the objective function (18) under the equivalence transformations $G \leftrightarrow GP$ of the system (4).

Proposition 4. *If the matrix* $P = \begin{bmatrix} p_1 & \dots & p_m \end{bmatrix}$ *in the condition of the Theorem 1 is nonsingular* ($m = N-n$)*, then the deviation matrix* $\mathscr{R}$ (26) *does not depend on* P.

6. Two special cases

Substituting $P = I_{N-n}$ in (22), we obtain the objective function of the variational identification method (VM) [4, 5] ($Y = V$), closely related to the Structured Total Least Squares problem [5, 14]. In turn, for $P = \begin{bmatrix} e_1 & e_{n+2} & \dots & e_M \end{bmatrix}$, $M = \left[\frac{N}{n+1}\right]$, where e_i denotes the i-th column of the identity matrix I_{Nn}, the expression (22) is the objective function of the classical orthogonal regression (OR, or the Total Least Squares problem) [1, 2, 3]. The choice of other values of P leads to a family of methods that are "intermediate" between OR and VM, differ in computational costs and numerical characteristics of the deviation ellipsoids. By the type of objective functions (18), (22), all these methods belong to the class of "orthoregressive" [5, 15]. Comparative analysis of particular cases and computational experiments are beyond the scope of this article.

Conclusion

Proceeding from the idea of M. Fliess about replacing a linear differential equation with an equivalent system of integral equations, it turns out to be possible to construct a whole family of operator-orthoregressive (ORR) methods for identifying the coefficients of linear difference equations, consistent under additive perturbations in measurements; the family of operator-orthoregressive methods includes the well-known orthogonal regression (TLS) and variational identification (STLS) methods as special cases. The conditions of identifiability are also obtained and quantitative indicators of local stability are constructed in terms of numerical characteristics of the ellipsoids of deviations of the OOR estimates.

This work was supported by the Russian Foundation for Basic Research (project No. 19-01-00754).

Appendix

Proof of Theorem 1. We establish an auxiliary statement.

Proposition 5. *Let* $p \in \mathbb{R}^{N-n}$*, matrices* G_α *and* E_j *be as defined in* (4), (5) *and* $\varphi_j \doteq E_j p$. *Then*

$$p^{\mathrm{T}} G_\alpha^{\mathrm{T}} x = 0 \quad \Leftrightarrow \quad \begin{bmatrix} \varphi_0^{\mathrm{T}} x & \dots & \varphi_n^{\mathrm{T}} x \end{bmatrix} \alpha = 0.$$

The proof follows from the definitions of (6), (8) and the equalities

$$p^{\mathrm{T}} G_\alpha^{\mathrm{T}} x = p^{\mathrm{T}} V \alpha = p^{\mathrm{T}} \begin{bmatrix} E_0^{\mathrm{T}} x & \dots & E_n^{\mathrm{T}} x \end{bmatrix} \alpha = \begin{bmatrix} \varphi_0^{\mathrm{T}} x & \dots & \varphi_n^{\mathrm{T}} x \end{bmatrix} \alpha.$$

Now we will prove the theorem. From the equations (3), (4) for an arbitrary matrix $P^{\mathrm{T}} \in \mathbb{R}^{m \times (N-n)}$ it follows that

$$P^{\mathrm{T}} G_\alpha^{\mathrm{T}} x = 0. \tag{27}$$

Conversely, take $m = N - n$ and suppose that the rows p_i^{T}, $i \in \overline{1, N-n}$ of the matrix P^{T} are linearly independent. Then the systems of equations (27) and (3), (4) are equivalent. According to Proposition 5, the system (27) is equivalent to

$$\left[\varphi_{i0}^{\mathrm{T}} x \quad \dots \quad \varphi_{in}^{\mathrm{T}} x\right] \alpha = 0, \quad i \in \overline{1, N-n},$$

which proves the theorem.

Proof of Lemma 1. The left-hand side of the equation $\det A^{\mathrm{T}} B = 0$ is a polynomial in the elements of the matrix $X \doteq A^{\mathrm{T}} B \in \mathbb{R}^{n \times n}$, so the equation defines a manifold in a matrix space $\mathbb{R}^{n \times n}$ such that, arbitrarily close to the matrix X, there is always a matrix $X_\varepsilon :\ \det X_\varepsilon \neq 0$. To complete the proof of the lemma, it suffices to prove the surjectivity of the mapping $A \mapsto X$, which follows from the completeness of the rank of the matrices A, B: $\forall X_\varepsilon\, \exists A_\varepsilon :\ A_\varepsilon^{\mathrm{T}} B = X_\varepsilon$. In the last sentence, one can swap the symbols A and B. This proves the lemma.

Proof of Theorem 2. For $Y\alpha = 0$ to be uniquely solved for a vector $\alpha \in \mathbb{R}^{n+1}$ with a fixed leading coefficient $\alpha_n = 1$, it is necessary and sufficient that the $\operatorname{rank} \overline{Y} = n$, where $\overline{Y}$ is the matrix composed of the first n columns of Y.

From Definitions (13), (6) and Proposition 5, we get $\overline{Y} = P^{\mathrm{T}} \overline{V} \in \mathbb{R}^{m \times n}$. Thus, to prove the theorem, it suffices to show that for almost all matrices P of full rank m the condition $\operatorname{rank} P^{\mathrm{T}} \overline{V} = n$ is equivalent to $\operatorname{rank} \overline{V}^{\mathrm{T}} \overline{V} = n$ and $m \geqslant n$.

Suppose that $\operatorname{rank} P^{\mathrm{T}} \overline{V} = n$. Hence $m \geqslant n$. Let $\overline{P}^{\mathrm{T}}$ denote a submatrix of n rows in P^{T}. By the hypothesis of the theorem, the rows in P^{T} are linearly independent, $\operatorname{rank} \overline{P}^{\mathrm{T}} = n$. Therefore, from $\operatorname{rank} \overline{P}^{\mathrm{T}}\, \overline{V} = n$ it follows that $\operatorname{rank} \overline{V} = n$ and $\operatorname{rank} \overline{V}^{\mathrm{T}} \overline{V} = n$.

Vice versa, let $\operatorname{rank} \overline{V}^{\mathrm{T}} \overline{V} = n$ and $m \geqslant n$. For almost any matrix P^{T} of full rank m, there is a submatrix $\overline{P}^{\mathrm{T}}$ of n linear independent rows in P^{T}. By Lemma 1, $\operatorname{rank} \overline{V} = n$ implies that $\operatorname{rank} \overline{P}^{\mathrm{T}}\, \overline{V} = n$. Since $\overline{P}^{\mathrm{T}}\, \overline{V}$ is a submatrix of rows in $P^{\mathrm{T}}\, \overline{V} \in \mathbb{R}^{m \times n}$, we have that $\operatorname{rank} P^{\mathrm{T}}\, \overline{V} = n$. This proves the theorem.

Proof of Corollary 1. The completeness of the solution x implies that $\operatorname{rank} \overline{V} = n$. Having $m \geqslant n$, to prove the corollary, it is enough to repeat the proof of the sufficiency part in Theorem 2.

Proof of Proposition 1. We use the following identity:

$$Y\alpha \equiv \left(I_{Lm} \otimes \alpha^{\mathrm{T}}\right) y \doteq \Gamma_\alpha^{\mathrm{T}} y.$$

Applying it to the matrix $Y \in \mathbb{R}^{Lm \times (n+1)}$ and vectors $\alpha \in \mathbb{R}^{n+1}$, $y = \operatorname{vect} Y \in \mathbb{R}^{Lm(n+1)}$, we can write (17) in the form

$$J(\alpha) = \min_{\Delta y \colon \Gamma_\alpha^{\mathrm{T}}(\check{y} - \Delta y) = 0} \|\Delta y\|_R^2 = \min_{y \colon \Gamma_\alpha^{\mathrm{T}} y = 0} \left(\check{y} - y\right)^{\mathrm{T}} R \left(\check{y} - y\right).$$

On the right-hand side of the equality, we find a point $\widehat{y}(\alpha)$ of minimum in y under the condition $\check{y} - y \in \operatorname{im} R$:

$$\widehat{y}(\alpha) \doteq \arg \min_{y \colon \Gamma_\alpha^{\mathrm{T}} y = 0} \left(\check{y} - y\right)^{\mathrm{T}} R \left(\check{y} - y\right) = \left[I_{Lm} - R^{+} \Gamma_\alpha \left(\Gamma_\alpha^{\mathrm{T}} R^{+} \Gamma_\alpha\right)^{-1} \Gamma_\alpha^{\mathrm{T}}\right] \check{y}.$$

Therefore,

$$J(\alpha) = \left(\check{y} - \widehat{y}(\alpha)\right)^{\mathrm{T}} R \left(\check{y} - \widehat{y}(\alpha)\right) = \check{y}^{\mathrm{T}} \Gamma_\alpha \left(\Gamma_\alpha^{\mathrm{T}} R^{+} \Gamma_\alpha\right)^{-1} \Gamma_\alpha^{\mathrm{T}} \check{y} = \\ = \alpha^{\mathrm{T}} \check{Y}^{\mathrm{T}} \left(\Gamma_\alpha^{\mathrm{T}} R^{+} \Gamma_\alpha\right)^{-1} \check{Y} \alpha,$$

which proves the proposition.

Proof of Proposition 2. The first formula of the proposition follows from the equalities (7), (11) and definition (20):

$$G_\alpha P = (\alpha_0 E_0 + \ldots + \alpha_n E_n) \begin{bmatrix} p_1 & \ldots & p_m \end{bmatrix} =$$
$$= \begin{bmatrix} (\alpha_0 E_0 + \ldots + \alpha_n E_n)\, p_1 & \ldots & (\alpha_0 E_0 + \ldots + \alpha_n E_n)\, p_m \end{bmatrix} =$$
$$= \begin{bmatrix} \backslash p_1 \backslash\, \alpha & \ldots & \backslash p_m \backslash\, \alpha \end{bmatrix} = \begin{bmatrix} \Sigma_1 \alpha & \ldots & \Sigma_m \alpha \end{bmatrix} = \Sigma \,(I_m \otimes \alpha)\,.$$

The second formula is obtained by substituting the first formula in the definitions of $\Gamma_\alpha \doteq I_{Lm} \otimes \alpha$ (19) and $R = I_L \otimes (\Sigma^{\mathrm{T}} \Sigma)^+$ (21).

Proof of Lemma 2. From Proposition 2, we have $\backslash \alpha \backslash\, P \doteq G_\alpha P = \Sigma\,(I_m \otimes \alpha)$. Next,

$$\Sigma\,(I_m \otimes \alpha) = \begin{bmatrix} \Sigma_1 & \ldots & \Sigma_m \end{bmatrix} (I_m \otimes \alpha) = \begin{bmatrix} \Sigma_1 \alpha & \ldots & \Sigma_m \alpha \end{bmatrix} \in \mathbb{R}^{N \times m}.$$

Let $f \doteq \begin{bmatrix} f_1 & \ldots & f_m \end{bmatrix}^{\mathrm{T}} \in \mathbb{R}^m$ be an arbitrary nonzero vector. Then

$$\begin{bmatrix} \Sigma_1 \alpha & \ldots & \Sigma_m \alpha \end{bmatrix} f = (f_1 \Sigma_1 + \ldots + f_m \Sigma_m)\, \alpha =$$
$$= (f_1 \backslash p_1 \backslash + \ldots + f_m \backslash p_m \backslash)\, \alpha = \backslash P f \backslash\, \alpha.$$

The columns of the matrix P are linearly independent; therefore, for any nonzero f, we have $Pf \neq 0$.

Claim. For any nonzero column vector g, the matrix $\backslash g \backslash$ has linearly independent columns.

Proof. Let g_* be the first nonzero element of the vector g, then $\backslash g \backslash = \begin{bmatrix} 0 & \ldots & 0 \\ g_* & & 0 \\ & \ddots & \\ * & & g_* \\ * & \ldots & * \end{bmatrix}$, which immediately implies the linear independence of the columns of $\backslash g \backslash$. □

By the claim, for any nonzero f and nonzero α, the product

$$\backslash P f \backslash\, \alpha = \Sigma\,(I_m \otimes \alpha)\, f$$

is also nonzero, which yields linear independence of the columns $\Sigma\,(I_m \otimes \alpha) = \backslash \alpha \backslash\, P$. So, the lemma is proved.

Proof of Corollary 2. Since the matrix P has full column rank, it follows from Lemma 2 that the columns of the matrix $G_\alpha P$ are linearly independent. According to Proposition 2, this means that the columns $\Sigma\,(I_m \otimes \alpha)$ are also linearly independent. Therefore, the matrix

$$\Gamma_\alpha^{\mathrm{T}} R^+ \Gamma_\alpha = (I_m \otimes \alpha^{\mathrm{T}})\, \Sigma^{\mathrm{T}} \Sigma\, (I_m \otimes \alpha)$$

is nonsingular.

The expression (22) for $J(\alpha)$ follows from (18), (19), and Proposition 2.

Proof of Proposition 3. We use the expressions for the derivatives obtained in [20, Lemmas 2, 3]:

$$\mathscr{R} = S^{\mathrm{T}} \Sigma^{\mathrm{T}} \Sigma S = (J''_{\theta\theta})^{-1} J''_{\theta y} \Sigma^{\mathrm{T}} \Sigma J''^{\mathrm{T}}_{\theta y} (J''_{\theta\theta})^{-1} =$$
$$= (D^{\mathrm{T}} Y^{\mathrm{T}} C_\theta Y D)^{-1} D^{\mathrm{T}} Y^{\mathrm{T}} C_\theta \Gamma_\theta^{\mathrm{T}} \Sigma^{\mathrm{T}} \Sigma \Gamma_\theta C_\theta Y D\, (D^{\mathrm{T}} Y^{\mathrm{T}} C_\theta Y D)^{-1}\,.$$

For $L = 1$, we have $C_\theta = (\Gamma_\theta^{\mathrm{T}} \Sigma^{\mathrm{T}} \Sigma \Gamma_\theta)^{-1}$ (see (19)). Then

$$\mathscr{R} = (D^{\mathrm{T}} Y^{\mathrm{T}} C_\theta Y D)^{-1} D^{\mathrm{T}} Y^{\mathrm{T}} C_\theta Y D\, (D^{\mathrm{T}} Y^{\mathrm{T}} C_\theta Y D)^{-1} =$$
$$= (D^{\mathrm{T}} Y^{\mathrm{T}} C_\theta Y D)^{-1}\,,$$

which proves the proposition.

Proof of Proposition 4. Taking into account Proposition 2, we have that

$$\mathscr{R} = (D^{\mathrm T}Y^{\mathrm T}C_\theta YD)^{-1} = \Big(D^{\mathrm T}Y^{\mathrm T}\left[P^{\mathrm T}G_\theta^{\mathrm T}G_\theta P\right]^{-1}YD\Big)^{-1} = \\ = \Big(D^{\mathrm T}Y^{\mathrm T}P^{-1}\left[G_\theta^{\mathrm T}G_\theta\right]^{-1}P^{-\mathrm T}YD\Big)^{-1}.$$

The following equalities are true:

$$P^{-\mathrm T}Y = P^{-\mathrm T}\begin{bmatrix}\varphi_{10}^{\mathrm T}x & \dots & \varphi_{1n}^{\mathrm T}x\\ \vdots & & \vdots\\ \varphi_{N-n,0}^{\mathrm T}x & \dots & \varphi_{N-n,n}^{\mathrm T}x\end{bmatrix} = P^{-\mathrm T}\begin{bmatrix}x^{\mathrm T}\varphi_{10} & \dots & x^{\mathrm T}\varphi_{1n}\\ \vdots & & \vdots\\ x^{\mathrm T}\varphi_{N-n,0} & \dots & x^{\mathrm T}\varphi_{N-n,n}\end{bmatrix} =$$

$$= P^{-\mathrm T}\begin{bmatrix}x^{\mathrm T}E_0p_1 & \dots & x^{\mathrm T}E_np_1\\ \vdots & & \vdots\\ x^{\mathrm T}E_0p_{N-n} & \dots & x^{\mathrm T}E_np_{N-n}\end{bmatrix} \doteq P^{-\mathrm T}\begin{bmatrix}x_{[0]}^{\mathrm T}p_1 & \dots & x_{[n]}^{\mathrm T}p_1\\ \vdots & & \vdots\\ x_{[0]}^{\mathrm T}p_{N-n} & \dots & x_{[n]}^{\mathrm T}p_{N-n}\end{bmatrix} =$$

$$= P^{-\mathrm T}\begin{bmatrix}p_1^{\mathrm T}x_{[0]} & \dots & p_1^{\mathrm T}x_{[n]}\\ \vdots & & \vdots\\ p_{N-n}^{\mathrm T}x_{[0]} & \dots & p_{N-n}^{\mathrm T}x_{[n]}\end{bmatrix} = \begin{bmatrix}p_1^{\mathrm T}\\ \vdots\\ p_{N-n}^{\mathrm T}\end{bmatrix}^{-1}\begin{bmatrix}p_1^{\mathrm T}x_{[0]} & \dots & p_1^{\mathrm T}x_{[n]}\\ \vdots & & \vdots\\ p_{N-n}^{\mathrm T}x_{[0]} & \dots & p_{N-n}^{\mathrm T}x_{[n]}\end{bmatrix} =$$

$$= P^{-\mathrm T}\begin{bmatrix}P^{\mathrm T}x_{[0]} & \dots & P^{\mathrm T}x_{[n]}\end{bmatrix} = \begin{bmatrix}x_{[0]} & \dots & x_{[n]}\end{bmatrix} = V.$$

As a result, $\mathscr{R} = \Big(D^{\mathrm T}V^{\mathrm T}\left[G_\theta^{\mathrm T}G_\theta\right]^{-1}VD\Big)^{-1}$. Thus, the deviation matrix $\mathscr{R}$ is independent of P if P is nonsingular.

References

[1] K. Pearson, *On lines and planes of closest fit to systems of points in space*, Phil. Mag., **6**:2 (1901), 559–572. JFM 32.0246.07

[2] H. Cramer, *Orthogonal mean square regression*, in *Mathematical methods of statistics (Ch. 23.6)*, Asia Publishing House, Bombay, NY, 1962. (Princeton, 1946. Zbl 0063.01014)

[3] G.H. Golub, C.F. Van Loan, *An analysis of the total least squares problem*, SIAM J. Numer. Anal., **17** (1980), 883–893. Zbl 0468.65011

[4] A.O. Egorshin, *Least squares method and the «fast» algorithms in variational problems of identification and filtration (VI method)* [in Russian], Avtometriya, **1988**:1 (1988), 30–41.

[5] A.A. Lomov, *Variation identification methods linear dynamic and the problem of local extremes* [in Russian], Large-scale Systems Control, **39** (2012), 53–94.

[6] Y.M. Bosinzon, I.Y. Falkovskii, *On a method for estimating parameters of time dependences through experimental data*, Autom. Remote Control, **55**:4 (1994), 502–505. Zbl 0843.93084

[7] S.N. Rogov, L.G. Sobolev, O.V. Khrutskii, *On some methods of smoothing and identification of experimental trends*, Autom. Remote Control, **66**:5 (2005), 799–809. Zbl 1080.62558

[8] B.O. Kachanov, *Symmetric Laplace transform and its application to parametric identification of linear systems*, Autom. Remote Control, **70**:8 (2009), 1309–1316. Zbl 1194.93046

[9] M. Fliess, H. Sira-Ramirez, *An algebraic framework for linear identification*, ESAIM, Control Optim. Calc. Var., **9** (2003), 151-168. Zbl 1063.93014

[10] R. Ushirobira, W. Perruquetti, M. Mboup, M. Fliess, *Algebraic parameter estimation of a biased sinusoidal waveform signal from noisy data*, 16th IFAC Symposium on System Identification, The International Federation of Automatic Control, Brussels, Belgium, 2012, 167–172.

[11] A.A. Lomov, *Operator-orthoregressive method for identifying coefficients of linear differential equations*, J. Math. Sci., **253**:3 (2021), 391–406.

[12] A.A. Lomov, *On quantitative a priori measures of identifiability of coefficients of linear dynamic systems*, J. Comput. Syst. Sci. Int., **50**:1 (2011), 1–13. Zbl 1267.93174

[13] I. Markovsky, S. van Huffel, *Overview of total least-squares methods*, Signal Process., **87**:10 (2007), 2283–2302. Zbl 1186.94229

[14] B. De Moor, *Structured total least squares and L_2 approximation problems*, Linear Algebra Appl., **188-189** (1993), 163–205. Zbl 0781.65028

[15] A.A. Lomov, *On consistency of a generalized orthoregressive parameter estimator for a linear dynamical system*, J. Appl. Ind. Mat., **8**:1 (2014), 86–91. Zbl 1340.93182

[16] A.A. Lomov, *On convergence of computational algorithms for a variational problem of identifying the coefficients of difference equations*, J. Appl. Industr. Math., **14**:3 (2020), 541–554.

[17] A.S. Householder, *On Prony's method of fitting exponential decay curves and multiple-hit survival curves*, Oak Ridge National Lab. Report ORNL-455, Oak Ridge, Tennessee, 1950.

[18] M.S. Mackisack, M.R. Osborne, G.K. Smyth, *A modified Prony algorithm for estimating sinusoidal frequencies*, J. Stat. Comput. Simulation, **49**:1-2 (1994), 111–124. Zbl 0833.62087

[19] I.G. Zedgenidze, *Designing an experiment to study multicomponent systems*, Nauka, Moscow, 1976.

[20] A.A. Lomov, *Local stability in the problem of identifying coefficients of a linear difference equation*, J. Math. Sci., **188**:4 (2013), 410–434. Zbl 1263.39017

Andrei Aleksandrovich Lomov
Sobolev Institute of Mathematics,
4, Koptyuga ave.,
Novosibirsk, 630090, Russia
Novosibirsk State University,
1, Pirogova str.,
Novosibirsk, 630090, Russia
Email address: lomov@math.nsc.ru, a.lomov@g.nsu.ru