DYNAMIC SYSTEM IDENTIFICATION BY USING SUBSPACE BASED METHODS AND ITS COMPARISON WITH CLASSICAL INPUT-OUTPUT APPROACH

Jakub Dokoupil
Doctoral Degree Programme (3), FEEC BUT
E-mail: xdokou05@stud.feec.vutbr.cz

Supervised by: Petr Pivoňka
E-mail: pivonka@feec.vutbr.cz

ABSTRACT

The thematic part of this article aims to compare the classic I/0 (Input-Output) approach and advanced subspace methods of dynamic system identification. Individual methods are compared in simulation with 8bits and 12bits digital-to-analog (D/A) and analog-to-digital (A/D) converters for various sampling periods.

1. INTRODUCTION

In this paper three methods for off-line parametric identification in open-loop are presented. The first two methods, which allow input-output identification (method of least squares and Gauss-Newton method), approach Hessian matrix inversion using singular value decomposition (SVD). Correlation and, in case of linear optimization, congruity was mathematically expressed in both methods. Third method combines in itself subspace identification method with the classical I/O approach and leads to state space representation process. In order to make the comparison correct, ARX model was chosen for the I/O description process.

2. INPUT-OUTPU IDENTIFICATION

Correspondence of model prediction with the output of identified process is a typical identification requirement. Optimal predictor of ARX model is

$$\hat{y}(k|k-1) = B(q)u(k) + \big(1 - A(q)\big)y(k) = \varphi(k+1)^T\theta \tag{1}$$

with polynomials of discrete transfer function $B(q)$ and $A(q)$. The aim is to find the parameters of model $\theta \in \mathbb{R}^n$ that minimize cost function $I(\theta)$ for measured input-output data from the process $\varphi(k) \in \mathbb{R}^n$

$$I(\theta) = \frac{1}{2}\sum_{i=0}^{p}\hat{e}(i)^2 = \frac{1}{2}E(\theta)^T E(\theta), \tag{2}$$

where p represents training data samples, $\hat{e}$ is prediction error and $E(\theta) \in \mathbb{R}^p$ is vector of prediction errors.

3. LINEAR LEAST SQUARES METHOD (LS)

Least squares method originated in connection with solving overdetermined set of equations. Its aim is to find the values of parameters $\theta \in \mathbb{R}^n$ that satisfy the system of equations for regressive data matrix $\phi \in \mathbb{R}^{p\times n}$ and vector for measured outputs $Y \in \mathbb{R}^p$ given

$$\phi\theta = Y. \tag{3}$$

For this reason solution exists only provided that the vector Y lies in the column space of ϕ. Optimum is found by minimizing the cost function

$$I(\theta) = \frac{1}{2}\|Y - \phi\theta\|_2^2. \tag{4}$$

Function $I(\theta)$ can be transcribed as

$$I(\theta) = \begin{bmatrix}1 & \theta^T\end{bmatrix} \underbrace{\begin{bmatrix}Y^TY & -Y^T\phi \\ -\phi^TY & \phi^T\phi\end{bmatrix}}_{M} \begin{bmatrix}1 \\ \theta\end{bmatrix}, \tag{5}$$

matrix M is factored using the Schur complement [1]

$$M = \begin{bmatrix}I_p & -\hat{\theta}^T \\ 0 & I_p\end{bmatrix}\begin{bmatrix}Y^TY - Y^T\phi\hat{\theta} & 0 \\ 0 & \phi^T\phi\end{bmatrix}\begin{bmatrix}I_p & 0 \\ -\hat{\theta} & I_p\end{bmatrix} \text{ with } \hat{\theta} = (\phi^T\phi)^{-1}\phi^TY. \tag{6}$$

Using factorization and its substitution into (5) we find that

$$I(\theta) = \left(Y^TY - Y^T\phi\hat{\theta}\right) + \left(\theta - \hat{\theta}\right)^T\phi^T\phi\left(\theta - \hat{\theta}\right)^T. \tag{7}$$

Only the second element of the numeric expression given above can be influenced by the choice of value θ. For this reason minimum $I(\theta)$ will be achieved when $\theta = \hat{\theta}$, thus

$$\theta = (\phi^T\phi)^{-1}\phi^TY. \tag{8}$$

The general solution of the equation (8) exists if the columns of matrix ϕ are linearly independent. However, if a matrix ϕ does not have a full column rank, solution is obtained using singular value decomposition (SVD), which is applied to matrix ϕ.

Singular value decomposition enables us to decompose the rectangular matrix $A \in \mathbb{R}^{m\times n}$ with rank r and given that $r < m$ and $r < n$ into product matrix in the form [1]

$$A = USV^T = \begin{bmatrix}U_1 & U_2\end{bmatrix}\begin{bmatrix}S_1 & 0 \\ 0 & 0\end{bmatrix}\begin{bmatrix}V_1^T \\ V_2^T\end{bmatrix} = U_1S_1V_1^T, \tag{9}$$

where $U_1 \in \mathbb{R}^{m\times r}$, $U_2 \in \mathbb{R}^{m\times(m-r)}$, $S \in \mathbb{R}^{r\times r}$, $V_1 \in \mathbb{R}^{n\times r}$, $V_2 \in \mathbb{R}^{n\times(n-r)}$.

From this relationship follows that matrix columns U_1, U_2, V_1 and V_2 provide orthogonal bases for four fundamental subspaces of the matrix: $\text{range}(A) = \text{range}(U_1)$, $\ker(A^T) = \text{range}(U_2)$, $\text{range}(A^T) = \text{range}(V_1)$, $\ker(A) = \text{range}(V_2)$.

Let SVD matrix ϕ according to (9) be

$$\phi = U_{1,\phi}S_{1,\phi}V_{1,\phi}^T, \tag{10}$$

then the direct method of least squares is

$$\theta = V_{1,\phi}\left(S_{1,\phi}\right)^{-1}U_{1,\phi}^TY. \tag{11}$$

4. GAUSS-NEWTON METHOD (GN)

Gauss-Newton method belongs to the family of gradient methods [2] which reduce the value of cost function $I(\theta)$ and which depend on parameters of model θ at each step of the calculation, thus $I(\theta_k) > I(\theta_{k+1})$. Restoration of parameters will be based on dependency

$$\theta_{k+1} = f\left(\theta_k,\ I(\theta_k),\ \frac{\partial}{\partial \theta_k} I(\theta_k),\ \ldots\right). \tag{12}$$

In case of penalization of prediction error by quadratic cost function, we can obtain solution by developing $E(\theta_k)$ into the Taylor series in the neighbourhood of θ_k

$$I(\theta_k + d_{k+1}) \approx \frac{1}{2}\left\| E(\theta_k) + \left(\frac{\partial}{\partial \theta_k^T} E(\theta_k)\right) d_{k+1} \right\|_2^2. \tag{13}$$

with added parameters $d_{k+1} = \theta_{k+1} - \theta_k$. Finding the optimal values of vector θ leads to the linear least squares problem and minimization of approximation we get

$$\theta_{k+1} = \theta_k - \left(J_{\hat{e}}^T(\theta_k) J_{\hat{e}}(\theta_k)\right)^{-1} J_{\hat{e}}^T(\theta_k) E(\theta_k) \tag{14}$$

with Jacobi matrix based on prediction errors $J_{\hat{e}}(\theta_k) \in \mathbb{R}^{p \times n}$

$$J_{\hat{e}}(\theta_k) = \frac{\partial E(\theta_k)}{\partial \theta_k^T} = \left[\frac{\partial \hat{e}(k|k-1)}{\partial \theta_k^T} \quad \frac{\partial \hat{e}(k-1|k-1)}{\partial \theta_k^T} \quad \ldots \quad \frac{\partial \hat{e}(k-p+1|k-1)}{\partial \theta_k^T}\right]^T.$$

Directional matrix $\left(J_{\hat{e}}^T(\theta_k) J_{\hat{e}}^T(\theta_k)\right)^{-1}$ significantly reduces the so-called zig-zag effect caused by heterogeneity of both partial and global minimum $I(\theta_k)$ due to possible asymmetry of the function $I(\theta_k)$ towards the main axes of parameters.

Singular value decomposition of Jacobi matrix $J_{\hat{e}}(\theta_k)$ is computed as follows

$$J_{\hat{e}}(\theta_k) = U_{1,J_{\hat{e}}}(\theta_k) S_{1,J_{\hat{e}}}(\theta_k) V_{1,J_{\hat{e}}}^T(\theta_k). \tag{15}$$

If the decomposition is substituted into the equation (14), utilizing orthogonal properties of matrices, minimum 2-norm is obtained (employing the SVD decomposition)

$$\theta_{k+1} = \theta_k - V_{1,J_{\hat{e}}}(\theta_k) \left(S_{1,J_{\hat{e}}}(\theta_k)\right)^{-1} U_{1,J_{\hat{e}}}^T(\theta_k) E(\theta_k). \tag{16}$$

For quadratic cost function and ARX model equation represents exact but not approximate solution according to the Taylor series. Under these conditions, the Gauss-Newton method can be regarded as an iterative version of the direct method of least squares.

5. DETERMINISTIC SYSTÉM IDENTIFICATION USING SUBSPACE METHOD

Subspace method identification is formulated for linear time-invariant deterministic discrete system in the form of

$$x(k+1) = Ax(k) + Bu(k), \tag{17}$$

$$y(k) = Cx(k) + Du(k), \tag{18}$$

where $x(k) \in \mathbb{R}^n$, $u(k) \in \mathbb{R}^m$ a $y(k) \in \mathbb{R}^l$.

The aim is to identify the system of matrices (A, B, C, D) as well as the initial state vector $x(0)$ of the minimum (reachable and observable) system, based on the final number of input-output data. The state representation of the system is not unique. Various realizations

of the system that maintain identical relation between input $u(k)$ and output $y(k)$ exist: input-output behaviour. We can thus transform the state vector $x(k)$ into $x_T(k)$

$$x_T(k) = T^{-1}x(k). \tag{19}$$

where T is regular matrix that is called state transformation. System corresponding to transformed state is given

$$x_T(k+1) = A_T x_T(k) + B_T u(k), \tag{20}$$

$$y(k) = C_T x_T(k) + D_T u(k), \tag{21}$$

where $A_T = T^{-1}AT$, $B_T = T^{-1}B$, $C_T = CT$, $D_T = D$.

State of system $x(k)$, provided that the initial state is $x(0)$, is at the time of k defined as

$$x(k) = A^k x(0) + \sum_{i=0}^{k-1} A^{k-i-1}Bu(i). \tag{22}$$

With regard to the equation (18) the relation between input data batch $\{u(k)\}_{k=0}^{s}$ and output data batch $\{y(k)\}_{k=0}^{s}$ can be specific

$$\underbrace{\begin{bmatrix} y(0) & \cdots & y(N-1) \\ \vdots & \ddots & \vdots \\ y(s-1) & \dots & y(N+s-2) \end{bmatrix}}_{Y_{0,s,N}} = \underbrace{\begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{s-1} \end{bmatrix}}_{\mathcal{O}_s} \underbrace{[x(0) \quad \dots \quad x(N-1)]}_{X_{0,N}}$$

$$+ \underbrace{\begin{bmatrix} D & 0 & \cdots & 0 \\ CB & D & \cdots & 0 \\ CAB & CB & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{s-2}B & CA^{s-3}B & \cdots & D \end{bmatrix}}_{\mathcal{T}_s} \underbrace{\begin{bmatrix} u(0) & \cdots & u(N-1) \\ \vdots & \ddots & \vdots \\ u(s-1) & \dots & u(N+s-2) \end{bmatrix}}_{U_{0,s,N}}. \tag{23}$$

In order to calculate the system matrix, properties of SVD will be used. The requirement is to find the column space of the extended observability matrix $\mathcal{O}_s$. Because matrix $\mathcal{T}_s$ is not known we cannot obtain the estimate of the column space by subtracting $\mathcal{T}_s U_{0,s,N}$ from $Y_{0,s,N}$ and by subsequent application of singular value decomposition. Instead, a formula using the estimate $\mathcal{T}_s$ [1] obtained by minimization will be used

$$\min_{T_s} \left\| Y_{0,s,N} - \mathcal{T}_s U_{0,s,N} \right\|_2^2. \tag{24}$$

Putting the result back into the equation (23) we get

$$Y_{0,s,N} - \hat{\mathcal{T}}_s U_{0,s,N} = Y_{0,s,N}\left(I_N - U_{0,s,N}^T\left(U_{0,s,N}U_{0,s,N}^T\right)^{-1}U_{0,s,N}\right) = Y_{0,s,N}H_{U_{0,s,N}}^{\perp}. \tag{25}$$

Matrix $H_{U_{0,s,N}}^{\perp}$ is orthogonal projection into the column space $U_{0,s,N}$ and has property $U_{0,s,N}H_{U_{0,s,N}}^{\perp} = 0$. By multiplying the equation (23) by the matrix $H_{U_{0,s,N}}^{\perp}$, the input-to-output influence will be eliminated. For this reason matrix $Y_{0,s,N}H_{U_{0,s,N}}^{\perp}$ contains column space of the extended observability matrix $\mathcal{O}_s$.

By solving the problem of minimization of error prediction model, we get the remaining matrices B_T and D_T together with the initial state $x_T(0)$

$$\hat{y}(k) = C_T A_T^k x_T(0) + \left(\sum_{\tau}^{k-1} u(\tau)^T \otimes C_T A_T^{k-\tau-1}\right)\mathrm{vec}(B_T) + \left(u(\tau)^T \otimes I_l\right)\mathrm{vec}(D_T). \tag{26}$$

6. SIMULATION RESULTS

Impact of quantization effect of various levels influences the identification quality of individual methods depending on the sampling period T_s is shown in the simulation of the process with transfer function

$$G_S(s) = \frac{2}{(10s+1)(s+1)^2}. \quad (27)$$

A/D and D/A converters are connected directly on system input and output. Verification of model properties is done through cost function always for the same time interval of the form

$$R(\theta) = \frac{1}{p}\sum_1^p \hat{e}(i)^2. \quad (28)$$

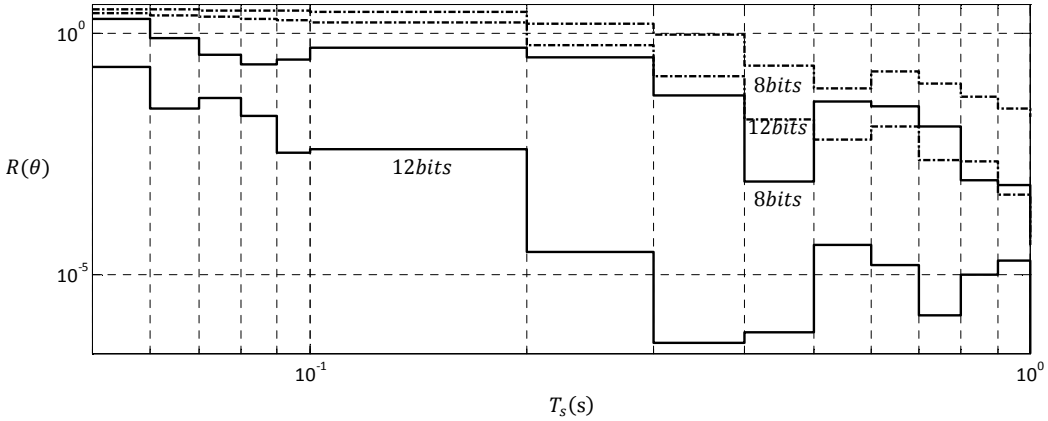

Fig. 1. Influence of quantization noise and sampling period on quality of identification using LS, GN (dot-line) and subspace based metod (solid line)

7. CONCLUSION

Results presented show that the subspace-based identification method achieves better properties for all levels of quantization noise in a defined spectrum of sampling periods. In both theoretical analysis and experimental simulation, equivalence of the direct method of least squares and the iterative Gauss-Newton method (or rather Newton method) was proved for quadratic cost function and for formulation of the prediction model in the form of linear regression. After meeting the conditions for solvability of the equation (3), both the method of least squares and Gauss-Newton method reach minimum of linear optimization problem in one step. Problem of Hessian matrix singularity was solved using SVD which ensures optimal solution even in numerically ill-conditioned cases. Simulation results were obtained in MATLAB/Simulink environment and individual algorithms can be implemented directly into PLC B&R.

REFERENCES

[1] Verhaegen, M., Verdult, V.: Filtering and system identification: a least squares approach, New York, Cambridge University Press 2007, ISBN 0-521-87512-9

[2] Nelles, Oliver.: Nonlinear system identification, New York, Springer-Verlag Berlin Heidelberg 2001, ISBN 3-540-67369-5