SOME INNOVATIONS TO THE MULTIPLE INPUT MULTIPLE OUTPUT IDENTIFICATION AND CONTROL

By

L. KEVICZKY and CS. BÁNYÁSZ*

Department of Automation, Technical University Budapest

(Received March 7, 1975)

Presented by Prof. Dr. F. CSÁKI

Introduction

The general use of computer technique leads to the wide application of discrete-time models in the system identification. Today the so-called ÅSTRÖM model and the maximum likelihood (ML) method are almost exclusive means of the parametric identification of linear dynamic systems [1] while the other methods — e.g., the generalized least squares (GLS), least squares (LS), priori knowledge fitting (PKF), first and second extended matrix methods, instrumental variable method, etc. — are only applied in special cases of the above model [1], [4].

Nowadays the theory related to the identification of linear, dynamic systems can be considered as more or less closed for single input single output (SISO) systems. Further efforts are, of course, directed to the parametric identification of multiple input multiple output (MIMO) systems. Certain authors suggest the use of state space representation for identification purposes since most important methods of the modern control theory involve the state space description of the system. Others prefer to identify the parameter matrices of a vector difference equation since this approach permits an easier generalization of the methods elaborated for SISO systems.

In this paper this latter approach is followed. The multiple output versions of LS, GLS, ML and PKF methods are given for special forms of vector difference equation in accordance with the MIMO generalization of the ÅSTRÖM model, and the multivariable form of minimum variance regulator is also presented.

The applied MIMO system model

A MIMO linear, discrete-time system can be described by the vector difference equation [1], [6]:

$$\mathbf{y}(t) = \sum_{i=0}^{n} \mathbf{B}_i \mathbf{u}(t-i) - \sum_{i=1}^{n} \mathbf{A}_i \mathbf{y}(t-i) + \sum_{i=1}^{n} \mathbf{C}_i \mathbf{e}(t-i) + \mathbf{e}(t) \qquad (1)$$

* Computer and Automation Institute of the Hungarian Academy of Sciences

where the environmental noise is also considered as in the Åström model. Here $\mathbf{y}$ is an $(q \times 1)$vector of the outputs, $\mathbf{u}$ is an $(m \times 1)$ vector of inputs, $\mathbf{e}$ is a $(q \times 1)$ vector of so-called source noise causing the corrupting effects at the outputs. The covariance matrix of $\mathbf{e}(t)$ is assumed to be $\mathbf{\Lambda}$ for all t and vectors $\mathbf{e}(t)$ and $\mathbf{e}(t+j)$ belonging to different times to be uncorrelated. The $(q \times m)$ and $(q \times q)$ matrices $\mathbf{B}_i$ and $\mathbf{A}_i$ contain the parameters of the process. The matrices $\mathbf{C}_i$ of $(q \times q)$ dimension include the parameters of the noise model. In Eq. (1) n is the order of the vector difference equation (and none of the system!) and t means the discrete (integer) time of the process.

Introducing the matrix polynomials:

$$\left.\begin{aligned} \mathbf{A}(z^{-1}) &= \mathbf{I}_q + \mathbf{A}_1 z^{-1} + \ldots + \mathbf{A}_n z^{-n} = \mathbf{I}_q + \tilde{\mathbf{A}}(z^{-1}) \\ \mathbf{B}(z^{-1}) &= \mathbf{B}_0 + \mathbf{B}_1 z^{-1} + \ldots + \mathbf{B}_n z^{-n} \\ \mathbf{C}(z^{-1}) &= \mathbf{I}_q + \mathbf{C}_1 z^{-1} + \ldots + \mathbf{C}_n z^{-n} \end{aligned}\right\} \tag{2}$$

the system equation can be given in the following form:

$$\mathbf{A}(z^{-1})\mathbf{y}(t) = \mathbf{B}(z^{-1})\mathbf{u}(t) + \mathbf{C}(z^{-1})\mathbf{e}(t) \tag{3}$$

where $\mathbf{I}_q$ is the $(q \times q)$ unit matrix and z^{-1} is the backward shift operator.

Now let us introduce the parameter matrix $\mathbf{P}$ containing the parameters of the process:

$$\mathbf{P} = [\mathbf{B}_0, \mathbf{B}_1, \ldots, \mathbf{B}_n, \mathbf{A}_1, \ldots, \mathbf{A}_n] \tag{4}$$

and the so-called memory vector belonging to the t-th observation:

$$\mathbf{x}(t) = [\mathbf{u}^T(t), \mathbf{u}^T(t-1), \ldots, \mathbf{u}^T(t-n), -\mathbf{y}^T(t-1), \ldots, -\mathbf{y}^T(t-n)]^T \tag{5}$$

where T means the transposition. By the above notations the system equation (1) becomes the following form:

$$\mathbf{y}(t) = \mathbf{P}\mathbf{x}(t) + \mathbf{C}(z^{-1})\mathbf{e}(t) \tag{6}$$

whence it is well seen that this relationship is linear with respect to the process parameters.

The LS method for MIMO systems

The well-known condition of the applicability of the least-squares method is that the equation error $\mathbf{C}(z^{-1})\mathbf{e}(t)$ should be uncorrelated, but this is fulfilled only if $\mathbf{C}(z^{-1}) = \mathbf{I}_q$. Then the system equation has the form

$$\mathbf{y}(t) = \mathbf{P}\mathbf{x}(t) + \mathbf{e}(t). \tag{7}$$

If N simultaneously measured values of input and output vectors are available then the LS estimation of **P** can be performed. Let the matrices **Y**, **F**, **E** be:

$$\mathbf{Y} = [\mathbf{y}(1), \mathbf{y}(2), \ldots, \mathbf{y}(N)] \tag{8}$$

$$\mathbf{F} = [\mathbf{x}(1), \mathbf{x}(2), \ldots, \mathbf{x}(N)] \tag{9}$$

$$\mathbf{E} = [\mathbf{e}(1), \mathbf{e}(2), \ldots, \mathbf{e}(N)] \tag{10}$$

then the joint system equation with respect to N measurements is:

$$\mathbf{Y} = \mathbf{PF} + \mathbf{E}\,. \tag{11}$$

At the LS method the loss function to be minimized is

$$V = \frac{1}{2}[\mathrm{vec}\,(\mathbf{E})]^T (I_N \otimes \mathbf{\Lambda}^{-1})\,\mathrm{vec}\,(\mathbf{E}) = \min_{\mathbf{P}} = \min_{\mathrm{vec}\,(\mathbf{P})} = = \min_{\mathbf{p}} \tag{12}$$

where the index N refers to the size and $\otimes$ means the Kronecker matrix product [5], [7].

In order to make the determination of the minimum easier we apply the Kronecker matrix operation of arrangement into the vector, $\mathbf{p} = \mathrm{vec}\,(\mathbf{P})$ [5], [7]. The necessary condition of the extremum is: $\partial V/\partial \mathbf{p} = \mathbf{0}$, whence an explicit expression is got for the parameter estimation $\hat{\mathbf{p}}$:

$$\hat{\mathbf{p}} = [(\mathbf{F}\,\mathbf{F}^T)^{-1}\,\mathbf{F} \otimes \mathbf{I}_q]\,\mathbf{w} \tag{13}$$

where

$$\mathbf{w} = \mathrm{vec}\,(\mathbf{Y})\,. \tag{14}$$

Even if the special condition $\mathbf{\Lambda} = \lambda^2\,\mathbf{I}_q$ holds, (13) simplifies into:

$$\hat{\mathbf{p}} = (\mathbf{F}^T\,\mathbf{F})^{-1}\,\mathbf{F}^T\mathbf{w}\,. \tag{15}$$

Taking into account the Kronecker matrix product identities, on the basis of (13):

$$\hat{\mathbf{P}} = \mathbf{Y}\mathbf{F}^T\,(\mathbf{F}\mathbf{F}^T)^{-1}\,. \tag{16}$$

This latter parameter estimation is unbiased if $\mathbf{u}(t)$ and $\mathbf{e}(t)$ are uncorrelated and it is consistent if $\mathbf{u}(t)$ fulfils the conditions of persistently exciting [1], [4]. If $\mathbf{e}(t)$ has a normal distribution, then the estimation is a ML estimation, too.

Extension of GLS method for MIMO systems

The generalization of CLARKE's GLS method [3] for MIMO systems means that the validity of approach $\mathbf{C}(z^{-1}) = \mathbf{H}^{-1}(z^{-1})$ is assumed in system Equation (3). Then the equation can be written in the form

$$\mathbf{H}(z^{-1})\,\mathbf{y}(t) = \mathbf{H}(z^{-1})\,\mathbf{B}(z^{-1})\,\mathbf{u}(t) - \mathbf{H}(z^{-1})\,\widetilde{\mathbf{A}}(z^{-1})\,\mathbf{y}(t) + \mathbf{e}(t)\,. \tag{17}$$

Considering, that

$$\left.\begin{aligned} \mathbf{H}(z^{-1})\,\mathbf{y}(t) &= [\mathbf{y}^T(t) \otimes \mathbf{H}(z^{-1})]\ \mathrm{vec}\ (\mathbf{I}_q) \\ \mathbf{H}(z^{-1})\,\mathbf{B}_i\mathbf{u}(t-i) &= [\mathbf{u}^T(t-i) \otimes \mathbf{H}(z^{-1})]\ \mathrm{vec}\ (\mathbf{B}_i) \\ \mathbf{H}(z^{-1})\mathbf{A}_i\,\mathbf{y}(t-i) &= [\mathbf{y}^T(t-i) \otimes \mathbf{H}(z^{-1})]\ \mathrm{vec}\ (\mathbf{A}_i) \end{aligned}\right\} \tag{18}$$

and introducing the following notations for the filtered values in the GLS method:

$$\left.\begin{aligned} \mathbf{Y}_F(t-i) &= \mathbf{y}^T(t-i) \otimes \mathbf{H}(z^{-1}) \\ \mathbf{U}_F(t-i) &= \mathbf{u}^T(t-i) \otimes \mathbf{H}(z^{-1}) \\ \mathbf{y}_F(t) &= [\mathbf{y}(t^T) \otimes \mathbf{H}(z^{-1})]\ \mathrm{vec}\ (\mathbf{I}_q) \end{aligned}\right\} \tag{19}$$

the system equation (17) becomes

$$\mathbf{y}_F(t) = \mathbf{X}_F(t)\,\mathbf{p} + \mathbf{e}(t) \tag{20}$$

where $\mathbf{p} = \mathrm{vec}\ (\mathbf{P})$ as seen above and

$$\mathbf{X}_F(t) = [\mathbf{U}_F(t), \mathbf{U}_F(t-1), \ldots, \mathbf{U}_F(t-n), -\mathbf{Y}_F(t-1), \ldots, -\mathbf{Y}_F(t-n)]. \tag{21}$$

If N measurements are available, then

$$\left.\begin{aligned} \mathbf{w}_F &= [\mathbf{y}_F^T(1), \mathbf{y}_F^T(2), \ldots, \mathbf{y}_F^T(N)]^T \\ \mathbf{F}_F &= [\mathbf{X}_F^T(1), \mathbf{X}_F^T(2), \ldots, \mathbf{X}_F^T(N)]^T \\ \mathbf{d} &= [\mathbf{e}^T(1), \mathbf{e}^T(2), \ldots, \mathbf{e}^T(N)]^T \end{aligned}\right\} \tag{22}$$

where

$$\mathbf{w}_F = \mathbf{F}_F\,\mathbf{p} + \mathbf{d}. \tag{23}$$

Using the filter equations (19), the solution is expected from the above discussed LS estimation and the parameter estimation is:

$$\hat{\mathbf{p}} = (\mathbf{F}_F^T\mathbf{F}_F)^{-1}\,\mathbf{F}_F^T\,\mathbf{w} \tag{24}$$

according to (15).

To apply Eq. (19) $\mathbf{H}(z^{-1})$ has to be known. Assuming now $\mathbf{A}(z^{-1})$ and $\mathbf{B}(z^{-1})$ to be known and $\mathbf{H}(z^{-1})$ to be of structure:

$$\mathbf{H}(z^{-1}) = \mathbf{I}_q + \mathbf{H}_1 z^{-1} + \ldots + \mathbf{H}_k z^{-k}. \tag{25}$$

Let us rewrite the system equation (17):

$$\mathbf{H}(z^{-1})\,[\mathbf{A}(z^{-1})\mathbf{y}(t) - \mathbf{B}(z^{-1})\,\mathbf{u}(t)] = \mathbf{H}(z^{-1})\mathbf{r}(t) = \mathbf{e}(t) \tag{26}$$

where the variable

$$\mathbf{r}(t) = \mathbf{A}(z^{-1})\mathbf{y}(t) - \mathbf{B}(z^{-1})\mathbf{u}(t) = \mathbf{y}(t) - \mathbf{X}\hat{\mathbf{p}} \tag{27}$$

is introduced for the equation error. Considering the construction of $\mathbf{H}(z^{-1})$ the autoregressive $\mathbf{r}(t)$ will be:

$$\mathbf{r}(t) = -\mathbf{H}_1\mathbf{r}(t-1) - \ldots - \mathbf{H}_k\mathbf{r}(t-k) + \mathbf{e}(t). \tag{28}$$

By this equation the least-squares estimations of $\mathbf{H}_i$ are:

$$\hat{\mathbf{Q}} = \mathbf{R}\mathbf{G}^T(\mathbf{G}\mathbf{G}^T)^{-1} \tag{29}$$

similarly to (16), where

$$\left.\begin{aligned} \mathbf{Q} &= [\mathbf{H}_1, \mathbf{H}_2, \ldots, \mathbf{H}_k] \\ \mathbf{R} &= [\mathbf{r}(1), \mathbf{r}(2), \ldots, \mathbf{r}(N)] \\ \mathbf{G} &= [\mathbf{g}(1), \mathbf{g}(2), \ldots, \mathbf{g}(N)] \end{aligned}\right\} \tag{30}$$

and

$$\mathbf{g}(t) = [-\mathbf{r}^T(t-1), \ldots, -\mathbf{r}^T(t-k)]. \tag{31}$$

Since on the basis of (27) $\mathbf{r}(t)$ depends on only $\hat{\mathbf{p}}$, the well-known successive approximation technique of the GLS method can be developed easily. This means that first we assume $\mathbf{H}(z^{-1}) = \mathbf{I}_q$ and estimate $\hat{\mathbf{p}}$, then compute $\mathbf{R}$ and $\mathbf{G}$, and estimate $\hat{\mathbf{Q}}$. After this we follow the procedure but at the newer estimation of $\hat{\mathbf{p}}$ the estimated $\hat{\mathbf{H}}(z^{-1})$ is taken into account. Actually this iterative technique corresponds to a relaxation minimization on the space of parameters $\mathbf{A}_i$, $\mathbf{B}_i$ and $\mathbf{H}_i$, respectively.

It has to be emphasized that though the equations are very similar to the case of single output GLS method, but there are basic differences in the construction of the applied "whitening" filters (see equations in (19)). This difference derives from the fact that the multiplication by $H(z^{-1})$ is convertible in case of single output systems, while for multiple output systems the order of the multiplication by $\mathbf{H}(z^{-1})$ is fixed!

MIMO version of the ML estimation method

If in the system eqation (1) $\mathbf{e}(t)$ has normal distribution then the ML estimation can be used. Now the loglikelihood function is [6]:

$$L = -\frac{qN}{2} \ln 2\pi - \frac{N}{2} \ln |\mathbf{\Lambda}| - \frac{1}{2} \sum_{t=1}^{N} \mathbf{e}^T(t) \mathbf{\Lambda}^{-1} \mathbf{e}(t) \tag{32}$$

where $|\ldots|$ means the determinant. To get the ML estimation, L has to be maximized. Since $\hat{\mathbf{\Lambda}}$ also depends on the parameters $\hat{\mathbf{p}}$, therefore first $L(\hat{\mathbf{p}}, \hat{\mathbf{\Lambda}}(\hat{\mathbf{p}}))$ is maximized by $\hat{\mathbf{\Lambda}}$, or minimized by $\hat{\mathbf{\Lambda}}^{-1}$:

$$\frac{\partial L\,[\hat{\mathbf{p}}, \hat{\mathbf{\Lambda}}(\hat{\mathbf{p}})]}{\partial \hat{\mathbf{\Lambda}}^{-1}} = \frac{N}{2} \hat{\mathbf{\Lambda}} - \frac{1}{2} \mathbf{E}\mathbf{E}^T = \mathbf{0} \tag{33}$$

whence

$$\hat{\mathbf{\Lambda}} = \frac{1}{N} \mathbf{E}\mathbf{E}^T = \sum_{t=1}^{N} \mathbf{e}(t)\, \mathbf{e}(t)^T . \tag{34}$$

Substituting this latter expression into (32) L will directly depend on $\hat{\mathbf{p}}$:

$$L(\hat{\mathbf{p}}) = \frac{qN}{2} (\ln 2\pi + 1) - \frac{N}{2} \ln |\hat{\mathbf{\Lambda}}(\hat{\mathbf{p}})| . \tag{35}$$

Thus the maximization of $L(\hat{\mathbf{p}})$ by $\hat{\mathbf{p}}$ is equivalent to the minimization of the loss function

$$V(\hat{\mathbf{p}}) = |\, \hat{\mathbf{\Lambda}}(\hat{\mathbf{p}}) \,| \to \min_{\hat{\mathbf{p}}} \tag{36}$$

Considering now that

$$\mathbf{P} = [\mathbf{B}_0, \mathbf{B}_1, \ldots, \mathbf{B}_n, \mathbf{A}_1, \ldots, \mathbf{A}_n, \mathbf{C}_1, \ldots, \mathbf{C}_n] = [\mathbf{P}_b, \mathbf{P}_a, \mathbf{P}_c] \tag{37}$$

for a general system and $\mathbf{p} = \mathrm{vec}\,(\mathbf{P})$ contains the parameters $\mathbf{C}_i$, too, the loss function is, however, a nonlinear (higher than quadratic order) function of parameters, the minimization of V cannot be given by an explicit expression, but only by some iteration extremum seeking method. Effective versions of these kinds of methods need the computation of the gradient and the Hessian-matrix of second order derivatives of $V(\hat{\mathbf{p}})$. The gradient of the loss function is

$$\frac{\partial V(\hat{\mathbf{p}})}{\partial \hat{\mathbf{p}}} = \sum_{t=1}^{N} \frac{\partial \mathbf{e}^T(t)}{\partial \hat{\mathbf{p}}} \hat{\mathbf{\Lambda}}^{-1} \mathbf{e}(t) . \tag{38}$$

The elements of matrix of second order derivatives are:

$$\frac{\partial^2 V(\hat{\mathbf{p}})}{\partial \hat{p}_i\, \partial \hat{p}_j} = \sum_{t=1}^{N} \frac{\partial \mathbf{e}^T(t)}{\partial \hat{p}_i} \hat{\mathbf{\Lambda}}^{-1} \frac{\partial \mathbf{e}(t)}{\partial \hat{p}_j} + \sum_{t=1}^{N} \mathbf{e}^T(t) \hat{\mathbf{\Lambda}}^{-1} \frac{\partial^2 \mathbf{e}(t)}{\partial \hat{p}_i\, \partial \hat{p}_j} . \tag{39}$$

Let us introduce the following notations:

$$\left.\begin{array}{l} \mathbf{P}_b = [\mathbf{B}_0, \mathbf{B}_1, \ldots, \mathbf{B}_n]; \quad \mathrm{vec}\,(\mathbf{P}_b) = \mathbf{p}_b \\ \mathbf{P}_a = [\mathbf{A}_1, \ldots, \mathbf{A}_n] \quad ; \quad \mathrm{vec}\,(\mathbf{P}_a) = \mathbf{p}_a \\ \mathbf{P}_c = [\mathbf{C}_1, \ldots, \mathbf{C}_n] \quad ; \quad \mathrm{vec}\,(\mathbf{P}_c) = \mathbf{p}_c \end{array}\right\} \tag{40}$$

and

$$\left.\begin{array}{l} \mathbf{x}_u(t) = [-\mathbf{u}^T(t), -\mathbf{u}^T(t-1), \ldots, -\mathbf{u}^T(t-n)]^T \\ \mathbf{x}_y(t) = [\mathbf{y}^T(t-1), \ldots, \mathbf{y}^T(t-n)]^T \\ \mathbf{x}_e(t) = [\mathbf{e}^T(t-1), \ldots, \mathbf{e}^T(t-n)]^T . \end{array}\right\} \tag{41}$$

Then the system equation is:

$$\mathbf{C}(z^{-1})\mathbf{e}(t) = \mathbf{y}(t) + \mathbf{P}_a \mathbf{x}_y(t) + \mathbf{P}_b \mathbf{x}_u(t) \tag{42}$$

hence

$$\left.\begin{array}{l} \mathbf{C}(z^{-1}) \dfrac{\partial \mathbf{e}(t)}{\partial \mathbf{p}_b^T} = \mathbf{x}_u^T(t) \otimes \mathbf{I}_q \\ \mathbf{C}(z^{-1}) \dfrac{\partial \mathbf{e}(t)}{\partial \mathbf{p}_a^T} = \mathbf{x}_y^T(t) \otimes \mathbf{I}_q \\ \mathbf{C}(z^{-1}) \dfrac{\partial \mathbf{e}(t)}{\partial \mathbf{p}_c^T} = \mathbf{x}_e^T(t) \otimes \mathbf{I}_q \end{array}\right\} . \tag{43}$$

Thus

$$\mathbf{C}(z^{-1}) \frac{\partial \mathbf{e}(t)}{\partial \mathbf{p}^T} = [\mathbf{x}_u^T(t), \mathbf{x}_y^T(t), \mathbf{x}_e^T(t)] \otimes \mathbf{I}_q \tag{44}$$

where

$$\mathbf{p} = [\mathbf{p}_b^T, \mathbf{p}_a^T, \mathbf{p}_c^T]^T . \tag{45}$$

These expressions deduced above already yield the gradient and the approximate matrix of second order derivatives according to the first term of (39), since

$$\frac{\partial V(\hat{\mathbf{p}})}{\partial \hat{\mathbf{p}}\, \partial \hat{\mathbf{p}}^T} \cong \sum_{t=1}^{N} \frac{\partial \mathbf{e}^T(t)}{\partial \hat{\mathbf{p}}} \hat{\mathbf{\Lambda}}^{-1} \frac{\partial \mathbf{e}(t)}{\partial \hat{\mathbf{p}}^T} . \tag{46}$$

These relationships permit to construct e.g. an iterative method:

$$\hat{\mathbf{p}}_{k+1} = \hat{\mathbf{p}}_k - \left[\frac{\partial V(\hat{\mathbf{p}}_k)}{\partial \hat{\mathbf{p}}_k\, \partial \hat{\mathbf{p}}_k^T}\right]^{-1} \frac{\partial V(\hat{\mathbf{p}}_k)}{\partial \hat{\mathbf{p}}_k} . \tag{47}$$

According to experience, the complexity concomitant to the exact computation of second order derivatives (the second term of (39)) is not worth of a somewhat improved convergence. The relationships (43) produce the gradient by simple autoregressive filtering similar to the case of single output systems.

Generalization of the second extended matrix method for MIMO systems

Using the parameter matrix $\mathbf{P}$ in (37), the system equation (1) can be written as:

$$\mathbf{y}(t) = \mathbf{P}\mathbf{x}(t) + \mathbf{e}(t) \tag{48}$$

where

$$\begin{aligned}\mathbf{x}(t) = [\mathbf{u}^T(t), \ldots, \mathbf{u}^T(t-n), -\mathbf{y}^T(t-1), \ldots, -\mathbf{y}^T(t-n),\\ \mathbf{e}^T(t-1), \ldots, \mathbf{e}^T(t-n)]^T .\end{aligned} \tag{49}$$

For the total of N measurements, the system equation is:

$$\mathbf{Y} = \mathbf{P}\mathbf{F} + \mathbf{E} \tag{50}$$

where

$$\mathbf{F} = [\mathbf{x}(1), \ldots, \mathbf{x}(N)] . \tag{51}$$

Formally (49) is identical to (11) and now the LS method can be applied:

$$\hat{\mathbf{P}} = \mathbf{Y}\mathbf{F}^T(\mathbf{F}\mathbf{F}^T)^{-1} . \tag{52}$$

$\mathbf{x}(t)$ and $\mathbf{F}$ contain the values of the source noise $\mathbf{e}(t)$ which, of course, are not known a priori. To apply (51) a series of $\mathbf{e}(t)$ (e.g., zeros, which correspond to an LS estimation) have to be assumed, then by this $\hat{\mathbf{P}}$ the $\hat{\mathbf{e}}(t)$ values can be estimated:

$$\hat{\mathbf{e}}(t) = \mathbf{y}(t) - \hat{\mathbf{P}}\mathbf{x}(t) \tag{53}$$

After this, the iterative procedure goes on as for SISO systems [4].

PKF method for MIMO systems

Without going into details, the more important relationships will only be presented. At the identification, the priori knowledge fitting method (PKF) assumes input and output noise to be independent. The applied system equation is:

$$\mathbf{y}(t) = \sum_{i=0}^{n} \mathbf{B}_i \mathbf{u}(t-i) - \sum_{i=1}^{n} \mathbf{A}_i \mathbf{y}(t-i) + \mathbf{d}(t) \tag{54}$$

where

$$\mathbf{d}(t) = \mathbf{e}(t) + \sum_{i=1}^{n} \mathbf{A}_i \mathbf{e}(t-i) . \tag{55}$$

Here $\mathbf{e}(t)$ is the output noise assumed to be ergodic, zero mean, independent of input random noise. Since $\mathbf{d}(t)$ and $\mathbf{u}(t)$ are independent it can be written:

$$E\{\mathbf{d}(t)|\mathbf{u}(t), \ldots, \mathbf{u}(t-s)\} = \sum_{i=0}^{s} \mathbf{D}_i \mathbf{u}(t-i) = \mathbf{d}(t) + \boldsymbol{\epsilon}(t). \tag{56}$$

where $\boldsymbol{\epsilon}(t)$ is the vector of residual errors and $E\{\ldots\}$ means the expectation. Because of independence, the condition $\mathbf{D}_i \to \mathbf{O}$ must hold (Thally-principle) [8].

Comparing (54) and (56):

$$\sum_{i=0}^{s} \mathbf{D}_i \mathbf{u}(t-i) = \mathbf{y}(t) + \sum_{i=1}^{n} \mathbf{A}_i \mathbf{y}(t-i) - \sum_{i=0}^{n} \mathbf{B}_i \mathbf{u}(t-i) + \boldsymbol{\epsilon}(t). \tag{57}$$

For N measurement vectors (57) can be written as

$$\boldsymbol{\mathcal{E}} = \mathbf{PF} + \mathbf{RZ} - \mathbf{Y} \tag{58}$$

where

$$\left.\begin{aligned}
&\mathbf{P} = [\mathbf{B}_0, \mathbf{B}_1 \ldots, \mathbf{B}_n, \mathbf{A}_1, \ldots, \mathbf{A}_n] \\
&\boldsymbol{\mathcal{E}} = [\boldsymbol{\epsilon}(1), \ldots, \boldsymbol{\epsilon}(N)] \\
&\mathbf{R} = [\mathbf{D}_0, \mathbf{D}_1, \ldots, \mathbf{D}_s] \\
&\mathbf{Y} = [\mathbf{y}(1), \ldots, \mathbf{y}(N)] \\
&\mathbf{F} = [\mathbf{x}(1), \ldots, \mathbf{x}(N)] \\
&\mathbf{Z} = [\mathbf{v}(1), \ldots, \mathbf{v}(N)]
\end{aligned}\right\} \tag{59}$$

and

$$\left.\begin{aligned}
&\mathbf{x}(t) = [\mathbf{u}^T(t), \ldots, \mathbf{u}^T(t-n), -\mathbf{y}^T(t-1), \ldots, -\mathbf{y}^T(t-n)]^T \\
&\mathbf{v}(t) = [\mathbf{u}^T(t), \ldots, \mathbf{u}^T(t-s)]^T.
\end{aligned}\right\}$$

To determine the parameters of system equation (57) the condition

$$[\mathrm{vec}\,(\boldsymbol{\epsilon})]^T \mathrm{vec}\,(\boldsymbol{\epsilon}) \to \min_{\mathrm{vec}(\mathbf{R})} \tag{60}$$

must be ensured which, after the usual steps, leads to:

$$\hat{\mathbf{R}} = (\mathbf{Y} - \mathbf{PF})\,\mathbf{Z}^T (\mathbf{ZZ}^T)^{-1}. \tag{61}$$

Otherwise the best parameter estimation $\hat{\mathbf{P}}$ is obtained from the minimum condition

$$[\mathrm{vec}\,(\mathbf{R})]^T \mathrm{vec}\,(\mathbf{R}) \to \min_{\mathrm{vec}\,(\mathbf{P})} \tag{62}$$

for $\mathbf{D}_i \to \mathbf{0}$. Minimization gives the explicit expression:

$$\hat{\mathbf{P}} = \mathbf{YZ}\,(\mathbf{ZZ}^T)^{-1}\,(\mathbf{ZZ}^T)^{-1}\,\mathbf{ZF}^T\,[\mathbf{FZ}^T\,(\mathbf{ZZ}^T)^{-1}(\mathbf{ZZ}^T)^{-1}\,\mathbf{ZF}^T]^{-1}\,. \tag{63}$$

Minimum variance control of MIMO systems

Eq. (1) of the MIMO systems is suitable to extend the classical minimum variance (MV) regulator elaborated for single output systems [2] for the case of MIMO systems, assuming a dead-time d in the system (1), it will be of the form:

$$\mathbf{y}(t+d) = \mathbf{A}^{-1}(z^{-1})\mathbf{B}(z^{-1})\mathbf{u}(t) + \mathbf{A}^{-1}(z^{-1})\mathbf{C}(z^{-1})\mathbf{e}(t+d)\,. \tag{64}$$

According to the well-known separating theorem and using the decomposition

$$\mathbf{C}(z^{-1}) = \mathbf{A}(z^{-1})\mathbf{F}(z^{-1}) + z^{-d}\mathbf{G}(z^{-1}) \tag{65}$$

we get

$$\begin{aligned}\mathbf{y}(t+d) = \mathbf{F}(z^{-1})\mathbf{C}(z^{-1})\mathbf{B}(z^{-1})\mathbf{u}(t) + \mathbf{A}^{-1}(z^{-1})\mathbf{G}(z^{-1})\mathbf{C}^{-1}(z^{-1})\mathbf{A}(z^{-1})\mathbf{y}(t) + \\ + \mathbf{F}(z^{-1})\,\mathbf{e}(t+d) = \hat{\mathbf{y}}(t+d\,|\,t) + \mathbf{F}(z^{-1})\mathbf{e}(t+d)\end{aligned} \tag{66}$$

corresponding to the generalization of the prediction law given for single output systems. The MV regulator is obtained by equalizing the predicted value with the reference signal and computing the necessary input vector. In case of zero reference value the following MV control law governs the optimum input:

$$\mathbf{u}_0(t) = -\,\mathbf{B}^{-1}\,(z^{-1})\,\mathbf{G}(z^{-1})\,\mathbf{F}^{-1}\,(z^{-1})\,\mathbf{y}(t) \tag{67}$$

obtained after simple calculations from the condition

$$\hat{\mathbf{y}}(t+d\,|\,t) = \mathbf{0}\,. \tag{68}$$

Relationship (67) is formally similar to that elaborated for the SISO system. Here the main difficulty is due to the fixed order of the multiplication of matrix polynomials, impeding a simple extension of the self-tuning regulator for MIMO systems.

Summary

The paper discusses the generalization of discrete identification methods — elaborated for SISO systems — for MIMO systems. It also deals with the known MIMO identification methods via uniform mathematical description using the new methods developed by the authors. Algorithms of LS, GLS, ML, PKF, SEXM, IV methods for MIMO systems are considered. After a survey of identification methods and their comparison, the possibility of designing an MV controller for MIMO systems with constant parameters is shown. An explicit expression is given for the impulse response matrix of optimum controller.

References

1. Åström, K. J.—Eykhoff, P.: System Identification — A Survey. 2nd IFAC Symp. on Identification and Process Parameter Estimation. Prague 1970.
2. Åström, K. J.: Introduction to Stochastic Control Theory. Academic Press. New York 1970.
3. Clarke, D. W.: Generalized-Least-Squares Estimation of the Parameters of a Dynamic Model. 1st IFAC Symp. Identification in Automatic Control Systems, Prague 1967.
4. Eykhoff, P.: System Identification. John Wiley and Sons, New York 1974.
5. Neudecker, H.: Some Theorems on Matrix Differentiation with Special Reference to Kronecker Matrix Products. American Stat. Ass. Journal, 1969, September.
6. Rowe, I. H.: A Statistical Model for the Identification of Multivariable Stochastic Systems. Proc. IFAC Symp. on Multivariable Systems, Düsseldorf 1968.
7. Tracy, D. S.—Dwyer, P. S.: Multivariate Maxima and Minima with Matrix Derivatives. American Stat. Ass. Journal, 1969, December.
8. Valis, J.: On-line Identification of Multivariable Linear Systems of Unknown Structure from Input Data. 2nd IFAC Symp. on Identification and Process Parameter Estimation, Prague 1970.

Dr. László Keviczky
Dr. Csilla Bányász
H-1521 Budapest