Print ISSN: 1312-2622; On-line ISSN: 2367-5357
DOI: 10.247/itc-2013-0009

General Forms of a Class of Multivariable Regression Models

A. Efremov

Key Words: *MIMO model; linear parameterization; parameter matrix form; parameter vector form; stepwise regression.*

Abstract. *There are two possible general forms of multiple input multiple output (MIMO) regression models, which are either linear with respect to their parameters or non-linear, but in order to estimate their parameters, at a certain stage it could be assumed that they are linear. This is in fact the basic assumption behind the linear approach for parameters estimation. There are two possible representations of a MIMO model, which at a certain level could be fictitiously presented as linear functions of its parameters. One representation is when the parameters are collected in a matrix and hence, the regressors are in a vector. The other possible case is the parameters to be in a vector, but the regressors at a given instant to be placed in a matrix. Both types of representations are considered in the paper. Their advantages and disadvantages are summarized and their applicability within the whole experimental modelling process is also discussed.*

Introduction

Nowadays mankind possesses a tremendous number of data, regarding many real-life phenomena. The availability of such source of information is the main premise for the widespread usage of experimental modelling. When incorporating more (appropriate) factors in the model, this leads to better ability of the models to represent the specifics of the investigated systems behaviour. Naturally, in fields like economics, sociology, finance, medicine, etc., the models are MIMO. Besides, the linear models with respect to their parameters are preferable (and applicable in many cases), because the theory used for their determination is well known and easy to apply. For instance, the iterative methods for numerical optimization, which are frequently used when the model is a non-linear function of the parameters, are avoided in the case of linear parameterized models. Furthermore, these models are proven to be efficient representations of systems with even strong non-linear input-output relations. The key point is to transform the model into a linear parameterized form, by finding appropriate (non-linear) functions of the initial factors and/or the model outputs [3,4].

In some cases even non-linear models with respect to their parameters could be considered as linear. Such an example is given below, where the model is ARMAX. In spite of the fact that the estimation process becomes iterative, in many cases this approach provides acceptable models with fewer calculations, compared to the traditional non-linear approach.

The considerations in the paper are focused on multivariable regression models, which are linear parameterized or non-linear parameterized, but could be assumed as linear functions of their parameters.

General Forms

When a SISO model is (assumed to be) linear parameterized, it can be written in the following general form

(1) $y_k = \varphi_k^T \theta + e_k$,

where y_k is the system output, e_k is the residual, containing the output variation, not accounted by the model, all factors (possibly transformed) are in the regression vector φ_k and all model parameters are arranged in the vector θ . This form is convenient, because all parameters are combined in a single vector, which has to be estimated based on the available data. Obviously the dimensions of both vectors are equal and their product is the predicted output (scalar) by the model, which is denoted below by $\hat{y}_k$. This process (according to (1)) is a linear function of the parameters.

In this paper the possible representations of the dynamic MIMO linear parameterized models are investigated. For more detailed explanation see [3].

Let the input and output values in the current (k -th) time instant be gathered in vectors and let the model have m inputs and ℓ outputs. Then the vectors are

$$u_k = [u_{1,k} \quad u_{2,k} \quad \dots \quad u_{m,k}]^T \ ;$$

$$y_k = [y_{1,k} \quad y_{2,k} \quad \dots \quad y_{\ell,k}]^T \ .$$

The multiplication between the factors and the parameters returns the vector $\hat{y}_k \in R^\ell$. With this notation, there are two possible ways to represent the multivariable linear parameterized model in a general form (*figure 1*). One way is the parameters to be gathered in the matrix

and the regressors – in the vector $\varphi_k \in R^z$ (is the number of the factors). This form is [5,6]

(2) $y_k = \Theta^T \varphi_k + e_k$.

The other way is the parameters to be placed in the vector $\theta \in R^p$ (the model has p parameters), but the

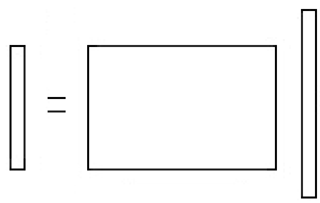

Figure 1. Model output $\hat{y}_k$ is a vector, so it could be represented as the product of a (factor/parameter) matrix and a (parameter/factor) vector

regressors at a given time instant to be arranged in the matrix $\Phi_k \in R^{\ell \times p}$. In this case the model becomes

(3) $y_k = \Phi_k \theta + e_k$.

The matrices and vectors are of appropriate structures – for more details see Sections 4 and 5. There are many realizations of each form, depending on how the parameters and the regressors are arranged. At first sight, both general forms must be equivalent, because, regardless of the right hand side, the result in both cases - (2) and (3), is the observed system output (here the residual is introduced in the expressions). But as it will be shown, (2) and (3) have specifics, which lead to different advantages, disadvantages and also to different applications.

MIMO ARX and ARMAX Models

When the input-output data is available (and it represents the dynamics of the investigated system), it is natural first to use an autoregressive model with exogenous input (ARX), which gives the relation between y_k and the values of the system inputs and outputs. After that, if an appropriate ARX structure is not found, another type of a regression model could be chosen. For example, the model could be extended by a noise filter. Probably the most widespread model from this group is ARMAX. It is an ARX model with a moving average (MA) noise filter. Alternatively, if the model type (ARX) remains the same, in order to improve the parameters' estimates accuracy, the instrumental variable method can be used.

ARX model has the form

(4) $\mathrm{A}(q^{-1})y_k = \mathrm{B}(q^{-1})u_k + e_k$.

Keeping in mind that $u_k \in R^m$ and $y_k \in R^\ell$, the polynomial matrices are $\mathrm{A}(q^{-1}) \in R_{na}^{\ell\times\ell}$ and $\mathrm{B}(q^{-1}) \in R_{nb}^{\ell\times m}$. By na and nb the maximum degrees of the polynomials in $\mathrm{A}(q^{-1})$ and $\mathrm{B}(q^{-1})$ are denoted. When the model (4) is MIMO, it can be considered as a system of ℓ equations. Each equation corresponds to a MISO sub-model, associated with a particular output.

If the statistical properties of the residual are not close to those of the white noise (in case of a colour residual, it will be denoted by $\tilde{e}_k$), the estimates are biased. To deal with this effect, the ARX structure could be extended with the following MA filter:

$\tilde{e}_k = \mathrm{C}(q^{-1})e_k$.

Here, at the stage of system identification, in addition to $\mathrm{A}(q^{-1})$ and $\mathrm{B}(q^{-1})$, the structure of $\mathrm{C}(q^{-1})$ and the parameters of its polynomials must be determined so that e_k to be white noise. The last requirement imposed on the residual provides unbiased parameters estimates. The practical reason is that white noise residual means that all deterministic input-output relations are accounted for by the model and the remaining output variation has only random components. As it was above mentioned, the extended ARX model with MA filter is named ARMAX. It has the form

(5) $\mathrm{A}(q^{-1})y_k = \mathrm{B}(q^{-1})u_k + \mathrm{C}(q^{-1})e_k$,

where $\mathrm{C}(q^{-1}) \in R_{nc}^{\ell\times\ell}$. Here nc is the maximum degree of the polynomials in $\mathrm{C}(q^{-1})$.

In the next two sections both ARX (linear parameterized) and ARMAX (a member of the non-linear parameterized models) are considered.

Parameter Matrix Form

It was mentioned so far, that there are many realizations of each general model form. Different ways of gathering the parameters in a matrix are considered below.

ARX Model

One realization of the parameter matrix form is the following. Both polynomial matrices in (4) can be represented as matrix polynomials:

(6) $\mathrm{A}(q^{-1}) = I_\ell + A_1 q^{-1} + \ldots + A_{na} q^{-na}$;

(7) $\mathrm{B}(q^{-1}) = 0_{\ell\times m} + B_1 q^{-1} + \ldots + B_{nb} q^{-nb}$,

where $A_i \in R^{\ell\times\ell}$ for $i = \overline{1, na}$ and $B_j \in R^{\ell\times m}$ for $j = \overline{1, nb}$ for consist of the parameters of all polynomials in $\mathrm{A}(q^{-1})$ and $\mathrm{B}(q^{-1})$, which are multiplied by y_{k-i} and u_{k-j} respectively. The intercept in (6) is chosen to be an identity matrix, because in each equation (a MISO sub-model) this provides the presence of only one of the outputs at the current time instant (the i-th equation contains only $y_{i,k}$). All other regressors are from previous time instants. This provides realistic model structures, simplifies the model determination and also its usage. The intercept in (7) is a zero matrix, because it is assumed that first y_k is measured, then the input u_k is determined (frequently as a function of y_k), and finally u_k is applied to the investigated system. Hence, the output y_k does not depend on the input $u_k = f(y_k)$. This is a standard case, when the model is used for prediction, control or other applications.

The matrix polynomials (6) and (7) may be used to obtain a specific realization of the general form (4), where all parameters are placed in a matrix.

Let (5) be rewritten as

(8) $\mathrm{A}(q^{-1}) = I_\ell + \tilde{\mathrm{A}}(q^{-1})$.

Then (4) can be formulated in the following way:

$$\begin{aligned} y_k &= -\tilde{\mathrm{A}}(q^{-1})y_k + \mathrm{B}(q^{-1})u_k + e_k \\ &= -A_1 y_{k-1} - \ldots - A_{na} y_{k-na} + B_1 u_{k-1} + \ldots + B_{nb} u_{k-nb} + e_k. \end{aligned}$$

Let the vector φ_k, containing all regressors be $\varphi_k = [-y_{k-1}^T \;\; \ldots \;\; -y_{k-na}^T \;\; u_{k-1}^T \;\; \ldots \;\; u_{k-nb}^T]^T$ and the parameter matrix Θ be

(9) $\Theta = [A_1 \;\; \ldots \;\; A_{na} \;\; B_1 \;\; \ldots \;\; B_{nb}]^T$.

Here the number of factors, necessary to predict each model output at a given time instant, is $z = \ell na + mnb$. Then the ARX model, represented in a general parameter matrix form becomes $y_k = \Theta^T\varphi_k + e_k$.

Different realizations of this form can be obtained by rearranging the regressors in φ_k (the structure of Θ should be changed as well, so that $\hat{y}_k = \Theta^T\varphi_k$ to be the same regardless of the different structure of φ_k).

From (8) it can be seen that when the parameters are gathered in matrix Θ, all polynomials in a polynomial matrix have to be from the same degree. For instance, let $na_{ij} = \deg(a_{ij}(q^{-1}))$ be known (and non-equal for all polynomials), where $a_{ij}(q^{-1})$ is the ij-th polynomial of $A(q^{-1})$. Then, in order to construct Θit is assumed that all polynomials degrees are $na = \max(na_{11}, na_{12}, \dots, na_{\ell\ell})$.

The same restriction is imposed on the polynomials degrees in $B(q^{-1})$. This is the main disadvantage of the parameter matrix form, which sometimes may lead to overfitted models and/or to numerical problems.

Another realization in the parameter matrix form could be obtained, if the columns of the polynomial matrices are considered ind ependently. Let $A(q^{-1})$ and $B(q^{-1})$ be presented as

$$A(q^{-1}) = [A_{.1}(q^{-1}) \quad A_{.2}(q^{-1}) \quad \dots \quad A_{.\ell}(q^{-1})],$$

$$B(q^{-1}) = [B_{.1}(q^{-1}) \quad B_{.2}(q^{-1}) \quad \dots \quad B_{.m}(q^{-1})],$$

where $A_{.i}(q^{-1})$ and $B_{.i}(q^{-1})$ are the columns of the matrices. Let also these polynomial vectors be written as vector polynomials, i.e.

$$A_{.i}(q^{-1}) = a_{i,0} + a_{i,1}q^{-1} + \dots + a_{i,na_i}q^{-na_i},$$

$$B_{.i}(q^{-1}) = b_{i,0} + b_{i,1}q^{-1} + \dots + b_{i,nb_i}q^{-nb_i}.$$

By na_i and nb_i the maximum degrees of the polynomials in the i-th columns of $A(q^{-1})$ and $B(q^{-1})$ are denoted. The elements of the vector $a_{i,0}$ are all zeroes except the i-th element which is equal to 1 and $b_{i,0}$ is a zero vector. Then the ARX model (4) becomes

$$A_{.1}(q^{-1})y_{1,k} + \dots + A_{.\ell}(q^{-1})y_{\ell,k} = B_{.1}(q^{-1})u_k + \dots + B_{.m}(q^{-1})u_k + e_k.$$

Taking in mind the above, the output y_k can be expressed as a function of the regressors in the following way:

$$\begin{aligned} y_k = &-a_{1,1}y_{1,k-1} - \dots - a_{1,na_1}y_{1,k-na_1} - \dots \\ &- a_{\ell,1}y_{1,k-1} - \dots - a_{\ell,na_1}y_{1,k-na_\ell} \\ &+ b_{1,1}u_{1,k-1} + \dots + b_{1,nb_1}u_{1,k-nb_1} + \dots \\ &+ b_{m,1}u_{m,k-1} - \dots - b_{m,nb_m}u_{m,k-nb_m}. \end{aligned}$$

The vectors a_{ij} and b_{ij} contain all polynomial parameters from i-th column of $A(q^{-1})$ and $B(q^{-1})$ respectively, which are multiplied by q^{-i}. Gathering all parameters in a matrix leads to (2), where

$$\begin{aligned} \varphi_k = [&-y_{1,k-1} \dots - y_{1,k-na_1} \quad \dots \quad -y_{\ell,k-1} \dots - y_{\ell,k-na_\ell} \\ &u_{1,k-1} \dots u_{1,k-nb_1} \quad \dots \quad u_{m,k-1} \dots u_{m,k-nb_m}]^T, \end{aligned}$$

$$\Theta = [a_{1,1} \dots a_{1,na_1} \dots a_{\ell,1} \dots a_{\ell,na_\ell} \; b_{1,1} \dots b_{1,nb_1} \dots b_{m,1} \dots b_{m,nb_m}]^T.$$

In this way the parameter matrix consists of $na_1 + na_2 + \dots + na_\ell$ vectors with ℓ components, containing the parameters in $A(q^{-1})$ and $nb_1 + nb_2 + \dots + nb_m$ vectors with ℓ components, containing the parameters in $B(q^{-1})$. Hence, the number of regressors is

$$z = \sum_{i=1}^{\ell} na_i + \sum_{i=1}^{m} nb_i$$

This realization is connected with more freedom for the model structure determination, since polynomials from different degrees may be placed in a polynomial matrix. But the restriction – all polynomials from a given column to have the same degree, remains.

ARMAX model

The first realization in the parameter matrix form is presented below for ARMAX model. The other form can be obtained in the same way as above described.

The polynomial matrix $C(q^{-1})$ can be represented as the following matrix polynomial

$$C(q^{-1}) = I_\ell + C_1q^{-1} + \dots + C_{nc}q^{-nc},$$

where $C_i \in R^{\ell \times \ell}$ for $i = 1, nc$ for consists of the parameters of all polynomials in $C(q^{-1})$, which are multiplied by e_{k-i}. After the same steps, applied for ARX model, one can arrive at (2), where the vector φ_k, containing all regressors, is $\varphi_k = [-y_{k-1}^T \dots - y_{k-na}^T \; u_{k-1}^T \dots u_{k-nb}^T \; e_{k-1}^T \dots e_{k-nc}^T]^T$ and the parameter matrix Θ is $\Theta = [A_1 \dots A_{na} \; B_1 \dots B_{nb} \; C_1 \dots C_{nc}]^T$.

The number of the factors is $z = \ell na + mnb + \ell nc$. Then the ARMAX model, represented in a parameter matrix form becomes $y_k = \Theta^T\varphi_k + e_k$.

Again different realizations of ARMAX model in the parameter matrix form can be obtained – if rearranging the regressors in φ_k and changing the structure of Θin the corresponding way.

Applications

From the discussions in the two previous subsections it follows that the representation in the parameter matrix is not appropriate for an accurate model structure determination. But if no apriori information is available about the polynomials degrees, (2) can be used for fast orientation in the model structure. This topic will be discussed later.

Another application of (2) is when restrictions are imposed on the model structure due to economic reasons. An example from finance industry and more precisely from the credit scoring activities is when a part of the factors is provided (on a corresponding price) by the credit bureaus. These factors are significantly more discriminative and from this point of view they are more desirable as factors in the model, compared to the other factors, like those, provided by the credit applicants. When a bank estimates how risky an applicant for a credit is, it may use a regression model to predict their behaviour. Normally the model requires some bureau factors. From this point of view, since each bureau factor costs money, it is reasonable to reduce the number of these factors in the model.

In this application, the model outputs can be the risk levels associated with different bank strategies, potential losses, etc. Therefore, if a bureau factor is introduced in the model to predict a specific output, then the same factor has to contribute to the prediction of the other outputs. If the regression model is in the form (2), then this restriction will be naturally accounted for in the process of model structure determination.

Another example for the application of (2), concerning the economic aspects is when building a control system. Here, each process included in the model is connected with an investment (especially in serial production). If a new input is incorporated in the model, then a corresponding actuator has to be provided and if an output is added, then a corresponding sensor is needed. In order to reduce the investment, the parameter matrix representation can be again used.

Parameter Vector Form

ARX Model

In order to transform the ARX model into the general form (3), it is convenient to start with the model, written as

$$(10)\quad y_k = -\tilde{A}(q^{-1})y_k + B(q^{-1})u_k + e_k,$$

where the matrix $\tilde{A}(q^{-1})$ is defined in (8). Let us introduce the block-diagonal matrices $Y_k \in R^{\ell\times\ell^2}$ and $U_k \in R^{\ell\times\ell m}$, which have ℓ and m blocks, which are y_k^T and u_k^T respectively. Let also $\underline{a}(q^{-1})$ and $\underline{b}(q^{-1})$ are the following polynomial vectors

$$\underline{a}(q^{-1}) = \mathrm{vec}(\tilde{A}^T(q^{-1})),$$

$$\underline{b}(q^{-1}) = \mathrm{vec}(B^T(q^{-1})).$$

By vec(.) the matrix vectorization is denoted. For matrix $M \in R^{m\times n}$ the resulting vector, after the vectorization, is

$$\mathrm{vec}(M) = [M_{.1}^T \;\; M_{.2}^T \;\; \dots \;\; M_{.n}^T]^T = [m_{11} \;\dots\; m_{1n} \;\;\dots\;\; m_{m1} \;\dots\; m_{mn}]^T \in R^{mn}.$$

Here $M_{.i}$ is the i-th column of M. Then the MIMO model (4) can be written as

$$y_k = -Y_k\underline{a}(q^{-1}) + U_k\underline{b}(q^{-1}) + e_k \;.$$

The multiplication between Y_k and $\underline{a}(q^{-1})$ leads to a vector with ℓ components. Actually, the i-th component is the scalar multiplication between the i-th row of $\tilde{A}(q^{-1})$ and y_k. It represents the overall influence of the previous values of the outputs on the current value of the i-th output. By analogy, the i-th element of $U_k\underline{b}(q^{-1})$ corresponds to the overall effect of the previous inputs on the i-th output. Indeed, because the polynomials in $\tilde{A}(q^{-1})$ and $B(q^{-1})$ (and hence in $\underline{a}(q^{-1})$ and $\underline{b}(q^{-1})$) do not have intercepts, all regressors in (4) are before k-th time instant.

In order to arrive at (3), it is necessary all parameters to be gathered in a vector θ. Also all regressors must be arranged in matrix Φ_k, with non-zero elements from the i-th row, which correspond to the regressors of the i-th MISO model. Moreover, they are placed in such an order that corresponds to the arrangement of the parameters in θ.

Again, as in the parameter matrix form, the elements of θ can be distributed in different ways. One case is the first elements of θ to be parameters of all polynomials in $\tilde{A}(q^{-1})$ (arranged sequentially row wise), followed by the polynomials' parameters in $B(q^{-1})$. According to this structure of θ, the regressors matrix Φ_k consists of two block diagonal sub-matrices. The first one is associated with the auto-regression, and the second one contains the previous values of the inputs.

Another example of the parameter vector form is the one, where the first $p_1 = \sum_{j=1}^{\ell} na_{1j} + \sum_{j=1}^{m} nb_{1j}$ elements of θ correspond to the polynomials from the first rows of $\tilde{A}(q^{-1})$ and $B(q^{-1})$. These elements of are all parameters of the first MISO model. The next $p_2 = \sum_{j=1}^{\ell} na_{2j} + \sum_{j=1}^{m} nb_{2j}$ elements of θ are the parameters of the second MISO model, etc. In this way Φ_k has a block diagonal structure, where the i-th block contains all p_i regressors of the i-th MISO model. The structure of Φ_k is presented in *figure 2*. This variant of the general representation is considered in more details below.

The polynomial matrices $\tilde{A}(q^{-1})$ and $B(q^{-1})$ can be written as

$$\tilde{A}(q^{-1}) = \begin{bmatrix} \tilde{a}_{11}(q^{-1}) & \dots & \tilde{a}_{1\ell}(q^{-1}) \\ \vdots & \ddots & \vdots \\ \tilde{a}_{\ell 1}(q^{-1}) & \dots & \tilde{a}_{\ell\ell}(q^{-1}) \end{bmatrix}$$

and

$$B(q^{-1}) = \begin{bmatrix} b_{11}(q^{-1}) & \dots & b_{m\ell}(q^{-1}) \\ \vdots & \ddots & \vdots \\ b_{\ell 1}(q^{-1}) & \dots & b_{m\ell}(q^{-1}) \end{bmatrix}$$

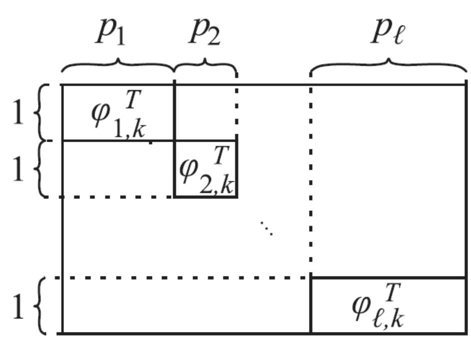

Figure 2. Structure of the regressors matrix Φ_k for ARX model

The j-th polynomial from the i-th row of matrix is $\widetilde{a}_{ij}(q^{-1}) = \widetilde{a}_{ij,1}q^{-1} + ... + \widetilde{a}_{ij,na_{ij}}q^{-na_{ij}}$ (for $i = \overline{1,\ell}$ and $j = \overline{1,\ell}$) and the polynomial $b_{ij}(q^{-1})$ is $b_{ij}(q^{-1}) = b_{ij,1}q^{-1} + ... + b_{ij,nb_{ij}}q^{-nb_{ij}}$, (for $i = \overline{1,\ell}$ and $j = \overline{1,m}$). Multiplying by the regressors, the i-th row of the matrix equation (4) becomes

$$y_{i,k} = -\widetilde{a}_{i1}(q^{-1})y_{1,k} - ... - \widetilde{a}_{i\ell}(q^{-1})y_{\ell,k} + b_{i1}(q^{-1})u_{1,k} + ... + b_{im}(q^{-1})u_{m,k} + e_{i,k}.$$

The terms on the right hand side can be written in the following vector form

$$\widetilde{a}_{ij}(q^{-1})y_{ij,k} = \underline{y}_{ij,k}^T \underline{a}_{ij},$$

$$b_{ij}(q^{-1})u_{ij,k} = \underline{u}_{ij,k}^T \underline{b}_{ij}.$$

The parameter vectors $\underline{a}_{ij} \in R^{na_{ij}}$ and $\underline{b}_{ij} \in R^{nb_{ij}}$, which are introduced above are

$$\underline{a}_{ij} = [\widetilde{a}_{ij,1} \quad \widetilde{a}_{ij,2} \quad ... \quad \widetilde{a}_{ij,na_{ij}}]^T$$

$$bij = [b_{ij,1} \quad b_{ij,2} \quad ... \quad b_{ij,nb_{ij}}]^T,$$

and the vectors $\underline{y}_{ij,k} \in R^{na_{ij}}$ and $\underline{u}_{ij,k} \in R^{nb_{ij}}$ containing the regressors are

$$\underline{y}_{ij,k} = [y_{j,k-1} \quad y_{j,k-2} \quad ... \quad y_{j,k-na_{ij}}]^T,$$

$$\underline{u}_{ij,k} = [u_{j,k-1} \quad u_{j,k-2} \quad ... \quad u_{j,k-nb_{ij}}]^T.$$

With this notation the i-th row of (2) becomes

$$y_{i,k} = -\underline{y}_{i1,k}^T \underline{a}_{i1} - ... - \underline{y}_{i\ell,k}^T \underline{a}_{i\ell} + \underline{u}_{i1,k}^T \underline{b}_{i1} + ... + \underline{u}_{im,k}^T \underline{b}_{im} + e_{i,k}$$

or briefly $y_{i,k} = \varphi_{i,k}^T \theta_i + e_{i,k}$.

Here all parameter vectors are gathered in the vector $\theta_i = [\underline{a}_{i1}^T \quad ... \quad \underline{a}_{i\ell}^T \quad \underline{b}_{i1}^T \quad ... \quad \underline{b}_{im}^T]^T \in R^{p_i}$, and the regressors are collected in the vector

$$\varphi_{i,k} = [-\underline{y}_{i1,k}^T \quad ... \quad -\underline{y}_{i\ell,k}^T \quad \underline{u}_{i1,k}^T \quad ... \quad \underline{u}_{im,k}^T]^T \in R^{p_i}.$$

Let us introduce the vector θ, which consists of the parameters of all MISO models. It has $p = \Sigma_i p_i$ elements and has the structure $\theta = [\theta_1^T \quad \theta_2^T \quad ... \quad \theta_\ell^T]^T$.

As the output vector is $y_k \in R^\ell$, then the regressors in the model have to be gathered in a matrix with ℓ rows and p columns. The non-zero elements of the i-th row must contain $\varphi_{i,k}$. The structure of the full MIMO ARX regressors matrix (shown in *figure 1*) at k-th time instant is

$$\Phi_k = \text{diag}(\varphi_{1,k}^T, \varphi_{2,k}^T, ..., \varphi_{\ell,k}^T).$$

It contains all regressors, necessary for the prediction of all model outputs. Finally the regression model (10) (and (4) respectively) in the parameter vector form is $y_k = \Phi_k \theta + e_k$.

When all model parameters are arranged in a vector, there is no restriction imposed on the polynomials degrees. As it is seen from the above considerations, the degrees (na_{ij} and nb_{ij}) of all polynomials in the model can be selected independently. Therefore, the representation in the parameter vector form can be used for accurate determination of the model structure.

ARMAX Model

Similar considerations to those in the previous subsection can be made for ARMAX model. Here the matrix $C(q^{-1})$ can be presented as

$$C(q^{-1}) = I + \begin{bmatrix} \widetilde{c}_{11}(q^{-1}) & ... & \widetilde{c}_{1\ell}(q^{-1}) \\ \vdots & \ddots & \vdots \\ \widetilde{c}_{\ell 1}(q^{-1}) & ... & \widetilde{c}_{\ell\ell}(q^{-1}) \end{bmatrix} = I + \widetilde{C}(q^{-1}),$$

where the j-th polynomial from the i-th row of matrix $\widetilde{C}(q^{-1})$ is $\widetilde{c}_{ij}(q^{-1}) = \widetilde{c}_{ij,1}q^{-1} + ... + \widetilde{c}_{ij,nc_{ij}}q^{-nc_{ij}}$ for $i, j = \overline{1,\ell}$. Thus the i-th row of the matrix equation (3) becomes

$$y_{i,k} = -\widetilde{a}_{i1}(q^{-1})y_{1,k} - ... - \widetilde{a}_{i\ell}(q^{-1})y_{\ell,k} + b_{i1}(q^{-1})u_{1,k} + ... + b_{im}(q^{-1})u_{m,k} + \widetilde{c}_{i1}(q^{-1})e_{1,k} - ... - \widetilde{c}_{i\ell}(q^{-1})e_{\ell,k} + e_{i,k},$$

where

$$\widetilde{c}_{ij}(q^{-1})e_{ij,k} = \underline{e}_{ij,k}^T \underline{c}_{ij}.$$

The parameter vectors $\underline{c}_{ij} \in R^{nc_{ij}}$ are

$$\underline{c}_{ij} = [\widetilde{c}_{ij,1} \quad \widetilde{c}_{ij,2} \quad ... \quad \widetilde{c}_{ij,nc_{ij}}]^T$$

and the vectors $\underline{e}_{ij,k} \in R^{nc_{ij}}$ are

$$\underline{e}_{ij,k} = [e_{j,k-1} \quad e_{j,k-2} \quad ... \quad e_{j,k-nc_{ij}}]^T.$$

The i-th row of (3) becomes

$$y_{i,k} = -\underline{y}_{i1,k}^T \underline{a}_{i1} - ... - \underline{y}_{i\ell,k}^T \underline{a}_{i\ell} + \underline{u}_{i1,k}^T \underline{b}_{i1} + ... + \underline{u}_{im,k}^T \underline{b}_{im} + \underline{e}_{i1,k}^T \underline{c}_{i1} - ... - \underline{e}_{i\ell,k}^T \underline{c}_{i\ell} + e_{i,k}$$

or briefly $y_{i,k} = \varphi_{i,k}^T \theta_i + e_{i,k}$, where $\theta_i = [\theta_{\text{ARX}i}^T \quad \theta_{\text{f}i}^T]^T$ and $\varphi_{i,k} = [\varphi_{\text{ARX}i,k}^T \quad \varphi_{\text{f}i,k}^T]^T$.

Here the ARX model parameters and the regressors are gathered in the vectors

$$\theta_{\text{ARX}i} = [\underline{a}_{i1}^T \quad ... \quad \underline{a}_{i\ell}^T \quad \underline{b}_{i1}^T \quad ... \quad \underline{b}_{im}^T]^T,$$

$$\varphi_{\text{ARX}i,k} = [-\underline{y}_{i1,k}^T \quad ... \quad -\underline{y}_{i\ell,k}^T \quad \underline{u}_{i1,k}^T \quad ... \quad \underline{u}_{im,k}^T]^T$$

and the filter parameters and regressors are gathered in

$$\theta_{\text{f}i} = [\underline{c}_{i1}^T \quad ... \quad \underline{c}_{i\ell}^T]^T,$$

$$\varphi_{\text{f}i,k} = [\underline{e}_{i1,k}^T \quad ... \quad -\underline{e}_{i\ell,k}^T]^T.$$

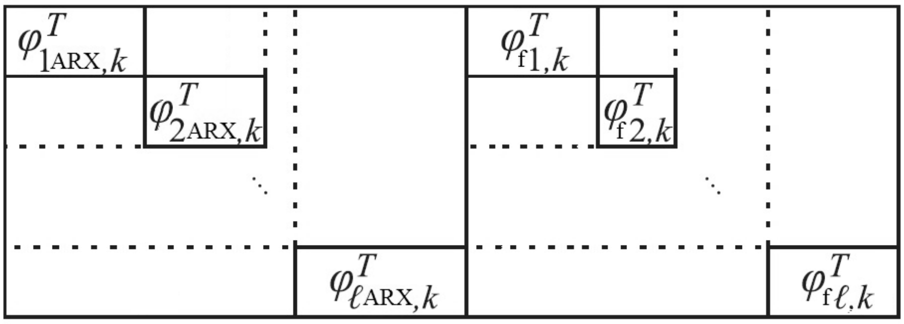

Figure 3. Structure of the regressors matrix Φ_k for ARMAX model

The reason to split the ARX and noise filter parts is that the first part depends directly on the observed input-output data, but the second part depends on the residual, which must be white noise. As the values of e_k depend on the specific model parameters, the noise part of the extended model has different properties from a modelling point of view.

Finally the vector *q*, which gathers all model parameters is $\theta = [\theta_1^T \;\; \theta_2^T \;\; \ldots \;\; \theta_\ell^T]^T$ and the regressors matrix (shown in *figure 3*) at *k*-th time instant is

$$\Phi_k = [\mathrm{diag}(\varphi^T_{\mathrm{ARX}1,k}, \; \ldots, \; \varphi^T_{\mathrm{ARX}\,\ell,k}) \quad \mathrm{diag}(\varphi^T_{\mathrm{f}1,k}, \; \ldots, \; \varphi^T_{\mathrm{f}\,\ell,k})] \, .$$

Finally, the regression model (5) in the parameter vector form becomes $y_k = \Phi_k \theta + e_k$.

Applications

As above mentioned, the representation in the parameter vector form can be used for accurate determination of the model structure, because there is no restriction imposed on the polynomials degrees.

Some dynamic systems, e.g., hypermarkets have many potential inputs [1]. Even after initial decomposition of the whole system, the resulting sub-systems may have hundreds or even thousands of potential inputs. Besides this, there are many possible cross relations within the multivariable structure, which have to be investigated when determining the market structure. In these cases, the number of competitive models, constructed by optimizing the structure parameters, may become very big. This leads to significant time for the overall model development process, especially when a parameter vector representation is used. For this reason, the general form (3) must be carefully used.

Factors Selection

Another point, which was not covered in the previous sections, is that the structure of each MISO model can be adjusted independently from the structures of the other sub-models [2]. On this basis efficient realizations of the methods for MIMO structure determination can be developed. Probably the most widespread method for this purpose is the stepwise regression. Its idea is shown in *figure 4*. Here the fact is used, that always the change of a particular structure parameter (time delay, polynomial degree, index of an entered input or output in the model), is actually a change in the set of factors, which are introduced in the model.

The stepwise method is iterative, where at each iteration two steps are consecutively applied. They are "forward selection" and "backward elimination". The result of the iterative process is a regression model, which accounts for a subset of factors, which (as a combination) are appropriate for explanation of the system output. In order to construct the model, the algorithm employs series of significance tests. The factors are added or removed one after another from the regression model. This process continues until no significant factors could be selected from the not entered ones and no insignificant factors could be eliminated from the model. The two thresholds in the stopping conditions, shown in *figure 3*, are SLE – the significant level for enter and SLS – the significant level to stay. With

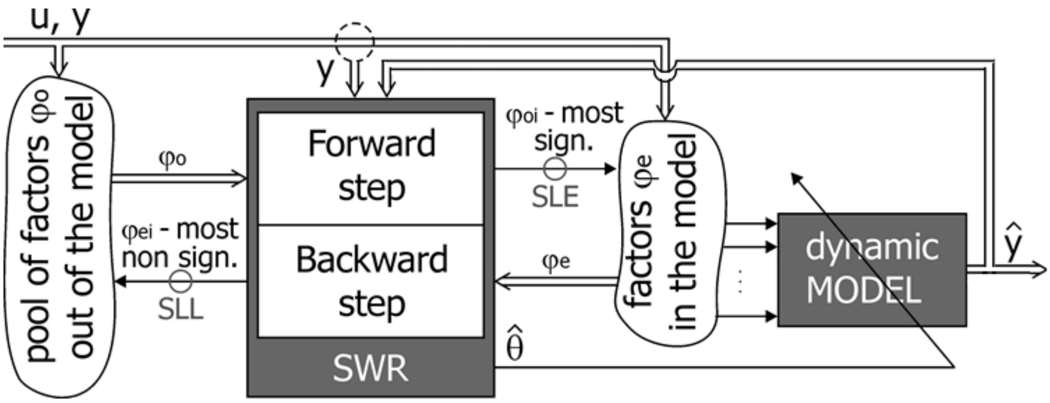

Figure 4. Stepwise regression

them the termination of the stepwise algorithm is controlled.

From the independence of the MISO sub-models, when the system has ℓ outputs, changes in the current MIMO model can be simultaneously applied. Moreover, at a given iteration of the stepwise method, some of the model changes (associated with certain outputs) could be a result of forward selection and at the same time other changes could be caused by backward elimination.

Conclusion

The two possible general representations of MIMO regression models in a (sometimes fictitious) linear parameterized form are presented. They are not equivalent and both can be used at certain steps during model development. In Section 4 it was shown that the representation with a parameter matrix leads to a constraint on the degrees of the model polynomials (they must be equal within a polynomial matrix or within the specific column of the matrix). This representation is applicable when the system dimension is very big and the model structure is unknown. In this case an appropriate solution at the stage of the model structure determination is first to use the model form with a parameter matrix for fast isolation of a subset of significant inputs and for initial orientation about the degrees of the model polynomials. After that the model structure can be further specified by using the parameter vector representation (where the above mentioned restriction is avoided). In this case the number of competitive models could be significantly reduced and the whole factors selection process becomes more efficient.

Another application of the parameter matrix representation is to take into account some non-statistical requirements (e.g. from economic perspective), introduced in the modelling problem. The constraint on the model structure, above mentioned, can be used to reduce the number of factors in the final model and thus to decrease the cost of model utilization.

Finally, keeping in mind the specific structures of (2) and (3), as mentioned in the previous section, there are ways to further increase the efficiency of the iterative algorithms for factors selection.

References

1. Efremov, A. Multivariate Time-Varying System Identification at Incomplete Information. Technical University of Sofia, Faculty of Automatics, Ph.D. 2008.
2. Efremov, A. System Identification Based on Stepwise Regression for Dynamic Market Representation. International Conference on Data Mining and Knowledge Engineering, Rome, Italy, 28–30 April 2010, 64, 2010, No. 2, 132-137.
3. Efremov, A. Multivariable System Identification. Monograph, Dar – RH, ISBN 978-954-9489-34-7, 2013.
4. Faraway, J. Practical Regression and ANOVA Using R. http://cran.r-project.org/doc/contrib/Faraway-PRA.pdf, 2002.
5. Van den Hof, P. M. J. Model Sets and Parameterizations for Identification of Multivariable Equation Error Models. – *Automatica*, 30 (3), 1994, 433-446.
6. Vuchkov, I. Identification. Sofia, IK Jurapel, 1996.

Manuscript received on 18.12.2013

Alexander Efremov *received his engineering degree in Systems and Control Engineering in 1999, and doctoral degree in production automation with application in the retail industry in 2008 from the Technical University of Sofia. From 2000 he started his career in the industry in Gallus Ltd as a control engineer, where he made a research on real-time identification for time-varying systems and adaptive control with application in the chemical industry. From 2000 to 2003 he was a software engineer in SkyGate BG Ltd (at present RaySat Ltd) where he worked in the field of satellite communication and flat antennas. His activities were related to modelling of multivariable nonlinear systems, Kalman filtering, interacting multiple model approach, etc. for satellite direction tracking. In the period 2004–2008 he was a researcher and scientific developer in Retail-Analytics Ltd. His work was focused on market systems identification and demand forecast for the retail sector of industry. Currently he works as a senior business analyst at Experian EAD, where his research is on modelling of human behavior and strategy optimization. The focus is on stepwise linear/non-linear regression, dynamic time-varying representations, decision trees, large scale sparse optimization, numerically stable realizations, Open CL and Hadoop technologies for parallel computations, etc. Meanwhile, he started his academic career in 2004 as an assistant at the University of Veliko Tarnovo and at the State University of Library Studies and Information Technologies. From 2005 to 2008 he was an assistant professor of the Department of Programming and Computer System Application of the University of Chemical Technology and Metallurgy. From 2006 he started his career at the Technical University of Sofia and now is a docent of the Industrial Department.*

Contacts:
Experian Ltd
Senior Business Analyst
tel: +359 896861315
e-mail: aefremov@gmail.com