Overlapping models merging and interconnection for large-scale model management

Pavel Trnka and Vladimír Havlena

Abstract—The application of advanced control methods to large-scale systems in variable industrial environment requires modeling and identification platform capable of keeping global model with description of its uncertainties, building global model from sub-systems, retrieval of sub-models with mutual consistency, model actualization from new sub-models or new data, etc. This article treats the problem of assembling global model for large-scale system from interconnected and possibly overlapping sub-models, i.e. there can be duplicity in the models. The quality of sub-models can also be different and is taken into account. The article presents two new results: merging of multiple models for the same system by using equivalent data and consistent combination of arbitrary connected models with parametric uncertainty into single model by using statistics of random vectors convolution.

I. INTRODUCTION

The appealing applications of advanced control algorithms to large-scale systems caused promising development of distributed algorithms in the recent years: distributed Kalman filtering [1], [2], distributed model predictive control [3], [4], distributed optimization [5], etc. Aside from this development stands a problem of identification and modeling with a focus on large-scale systems [6]. It is well known that (control oriented) model quality plays important role in the performance of distributed algorithms. Yet there are rather limited results and this field still waits for further exploration.

The application of advanced control to large-scale systems in variable industrial environment (power plants, water networks, chemical processes) requires modeling and identification platform capable of:

1) keeping global model with description of its uncertainty
2) consistent assembling of global model from possibly overlapping sub-models
3) retrieving arbitrary sub-models with selected complexity (model order)
4) retrieving hierarchical sub-models with guaranteed consistency between hierarchy levels
5) actualization of global model by new sub-models or newly acquired data
6) reflecting changes in system structure (valve closing, sub-system failure) without extensive model recomputations

This article deals mainly with the second problem of global model assembling and divides it into two separate problems: computing overall model of arbitrarily interconnected models with parametric uncertainties and merging of multiple models describing the same system.

To demonstrate these partial problems assume an interconnection of four systems in Figure 1 and knowledge of individual models for each system $M_1$, $M_2$, $M_3$, $M_4$ and model $M_{123}$ of $S_1$, $S_2$, $S_3$ in closed-loop. The overall model can be constructed by

1) computing model of $M_1$, $M_2$, $M_3$ in a closed loop (while respecting the individual model uncertainties)
2) merging the previous result with $M_{123}$ (both models describe the same dynamics)
3) combining the previous result in parallel connection with $M_4$

This can be generalized to a complex interconnection of overlapping models. The algorithm for combination of arbitrary connected models with uncertainty and the algorithm for merging of multiple models for single system are described in the following sections.

The selection of model class is also important as it is restricted by requirement to allow effective description of model uncertainty, it must be closed with respect to serial/parallel/feedback connection and merging of two models of the same system must be possible. With these requirements and with respect to simplicity ARX model class was selected.
Further assume a regressor with the following structure 

$$b$$

in parameters is not possible as the models can have a forward with apparent physical interpretation. However, the in "parameters domain" has to be replaced by merging in equivalent data ARX model by 

$$\text{resulting uncertainty in frequency domain cannot be easily transformed back to the uncertainty of ARX model parameters.}$$

Assume model parameters described by Normal distribution $\theta \sim N(\hat{\theta}, s^2 I)$, where $\hat{\theta}$ is a mean value, $P$ is a normalized covariance matrix and $s^2$ is a normalization constant (residual variance). Finding equivalent data means to find a regressor $Z$ and a RHS vector $b$ such that

$$Z\hat{\theta} = b, \quad Z^T Z = P^{-1}, \quad (2)$$

where $Z$ and $b$ have the correct Hankel structure induced by ARX model (1). Equivalent data can be generated in an arbitrary length. However, there is a lower limit for equivalence (2) to hold.

1) Minimum length of equivalent data: <The description of equivalent data minimum length will be provided in the final version.>

Having equivalent data for each model makes merging simple as equivalent data can be reordered into a regressor and a RHS vector of any structure. Reordering equivalent data of all merged models into the same structure of a target model allows to stack the regressors and the RHS vectors and to compute the final model parameters by the weighted least squares (Figure 2).

B. Generating equivalent data

The goal is to generate equivalent data set $\{u_1^{N_f}, y_1^{N_f}\}$ of length $N_f$ complying with the mean value $\hat{\theta}$ and the precision matrix $F = P^{-1}$ of original ARX model parameters.

$$Z\hat{\theta} = b, \quad Z^T Z = F,$$

where $Z$ and $b$ are assembled from $\{u_1^{N_f}\}$ and $\{y_1^{N_f}\}$ according to ARX model structure (1).

The first step is to find all feasible solutions for $Z\hat{\theta} = b$ while respecting $Z$ and $b$ structure. The structure can be enforced [9] by vec operator

$$\text{vec}(UVW) = (W^T \otimes U) \text{vec}(V), \quad (4)$$

leads to

$$\left( \left( \hat{\theta}^T \otimes I_{N_f} \right) N_Z - N_b \right) z = 0.$$  

Denoting $B = \ker \left( \left( \hat{\theta}^T \otimes I_{N_f} \right) N_Z - N_b \right)$, the solution is

$$z = Bu,$$

where $u$ is a new variable of lower dimension uniquely determining $Z$ as $\text{vec}(Z) = N_Z Bu$ and $b$ as $N_b Bu$. The next step is to find $u$ such that $Z^T Z = F$. Denoting Cholesky factor of positive definite precision matrix $F$ as $R$ gives

$$Z^T Z = R^T R,$$

$$Z \in \mathbb{R}^{N \times (n_a+n_b)} \quad R \in \mathbb{R}^{(n_a+n_b) \times (n_a+n_b)}.$$  

This implies that every feasible $Z$ can be written as

$$Z = TR,$$

where $T \in \mathbb{R}^{N \times (n_a+n_b)}$ is a transformation matrix with orthonormal columns. Assuming full rank of $F$ the transformation $T$ can be found as

$$T = ZR^{-1}, \quad (5)$$

which can be shown to be orthonormal

$$T^T T = R^{-T} R R^{-1} = I.$$  

1Notation $x_{i_1}^2$ stands for $(x_{i_1} \ldots x_{i_2})$. 

II. NOTATION

Assume ARX model [7] with a single input $u(t)$ and a single output $y(t)$

$$y(t) + a_1 y(t-1) + \cdots + a_{n_a} y(t-n_a) + e(t) = b_1 u(t-1) + \cdots + b_{n_b} u(t-n_b) + e(t),$$

with structural parameters $n_a, n_b$ and noise $e(t) \sim N(0, \sigma^2)$. Further assume a regressor $Z$ with $N$ rows and a RHS vector $b$ with the following structure

$$Z = \begin{pmatrix} y_{n_a} & \cdots & y_1 & u_{n_a} & \cdots & u_{n_a-n_b+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ -y_{n_a+N-1} & \cdots & -y_N & u_{n_a+N-1} & \cdots & u_{n_a-n_b+N} \end{pmatrix},$$

$$b = \begin{pmatrix} y_{n_a+1} \cdots y_{n_a+N} \end{pmatrix}^T, \quad (1)$$

and a vector of parameters

$$\theta = \begin{pmatrix} a_1 & \cdots & a_{n_a} \ \ b_1 & \cdots & b_{n_b} \end{pmatrix}^T.$$ 

III. MERGING TWO MODELS DESCRIBING THE SAME SYSTEM

It is obvious that direct merging of two ARX models in parameters is not possible as the models can have a different structure (different structural parameters). Merging in "parameters domain" has to be replaced by merging in time or frequency domain.

The merging in frequency domain [8] is quite straightforward with apparent physical interpretation. However, the resulting uncertainty in frequency domain cannot be easily transformed back to the uncertainty of ARX model parameters. On the other side, merging in time domain does not have this drawback – the merged model parameters are again described by a mean value and a covariance matrix.

A. Merging in time domain

Merging in time domain can be based on a description of ARX model by equivalent data, which is a set of input/output data resulting to the same estimate of model parameters mean value and covariance.

Assume model parameters described by Normal distribution $\theta \sim N(\hat{\theta}, \sigma^2 I)$, where $\hat{\theta}$ is a mean value, $P$ is a normalized covariance matrix and $\sigma^2$ is a normalization constant (residual variance). Finding equivalent data means to find a regressor $Z$ and a RHS vector $b$ such that

$$Z\hat{\theta} = b, \quad Z^T Z = P^{-1}, \quad (2)$$

where $Z$ and $b$ have the correct Hankel structure induced by ARX model (1). Equivalent data can be generated in an arbitrary length. However, there is a lower limit for equivalence (2) to hold.
and using again formula (4) yields

\[
\delta_\theta, \delta_{R} \text{ and solved for columns. This is unfortunately non-trivial problem of solving the equations.}
\]

\[
\text{(N}_b \text{B)} \begin{bmatrix} u \end{bmatrix} = \begin{bmatrix} \text{vec}(TR) \end{bmatrix} \]

Denoting

\[
\left( \frac{B_u}{B_T} \right) = \ker \left( \text{(N}_b \text{B)} \begin{bmatrix} u \end{bmatrix} = \begin{bmatrix} \text{vec}(TR) \end{bmatrix} \right),
\]

the solution in a new variable \( \mathbf{x} \) is

\[
\left( \frac{\mathbf{u}}{\text{vec}(\mathbf{T})} \right) = \left( \frac{B_u}{B_T} \right) \mathbf{x}.
\]

The last problem is to select \( \mathbf{x} \) such that \( \mathbf{T} \) has orthonormal columns. This is unfortunately non-trivial problem of solving a set of quadratic matrix equations. Introducing column selecting matrix \( C_i \) and using Matlab like notation

\[
\mathbf{T}(; i) = C_i \text{vec}(@) = C_i \mathbf{B}_T \mathbf{x},
\]

the problem can be formulated as

\[
\mathbf{T}(; i)^T \mathbf{T}(; j) = \mathbf{x}^T \mathbf{B}_T^T \mathbf{C}_i^T \mathbf{C}_j \mathbf{B}_T \mathbf{x} = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}
\]

and solved for \( \mathbf{x} \) by nonlinear least squares as

\[
\min_{\mathbf{x}} \sum_{i,j} \left( \mathbf{x}^T \mathbf{B}_T^T \mathbf{C}_i^T \mathbf{C}_j \mathbf{B}_T \mathbf{x} - \delta_{ij} \right)^2,
\]

where \( \delta_{ij} \) is Kronecker delta. Having solution \( \mathbf{x} \) the regressor \( \mathbf{Z} \) is obtained by reshaping vector \( \mathbf{N}_b \mathbf{B}_u \mathbf{x} \) into matrix and RHS vector \( \mathbf{b} \) is obtained as \( \mathbf{N}_b \mathbf{B}_u \mathbf{x} \).

**Algorithm 1** Equivalent data for ARX (FIR) model

**Inputs:**

\( N \) number of regressor rows
\( \theta, \mathbf{F} \) target mean value and normalized precision

**Outputs:**

\( \{u, y\}_1^N \) equivalent inputs/outputs

1. Find space of all unique solutions for \( \mathbf{Z} \mathbf{\hat{\theta}} = \mathbf{b} \) respecting \( \mathbf{Z} \) and \( \mathbf{b} \) structure as

\[
\mathbf{z} = \mathbf{B}_u \mathbf{b} = \ker \left( \left[ \begin{array}{c} \mathbf{\theta}^T \mathbf{\Theta} \mathbf{I}_N \end{array} \right] \mathbf{N}_z - \mathbf{N}_b \right),
\]

in a new variable \( \mathbf{u} \) using notation (3).

2. Restrict solutions in variable \( \mathbf{u} \) by target precision \( \mathbf{F} \) as

\[
\mathbf{u} = \mathbf{B}_u \mathbf{x}, \quad \left( \frac{\mathbf{B}_u}{\mathbf{B}_T} \right) = \ker \left( \left[ \begin{array}{c} \mathbf{N}_Z \mathbf{B} \mathbf{u} \end{array} \right] = \ker \left[ \left[ \mathbf{N}_Z \mathbf{B} \mathbf{u} \right] = \ker \left[ \left. \mathbf{F} \right] \mathbf{N}_z - \mathbf{N}_b \right) \right),
\]

in a new variable \( \mathbf{x} \), where \( \mathbf{R} \) is Cholesky factor of \( \mathbf{F} \).

3. Solution \( \mathbf{x} \) is obtained from non-linear least squares

\[
\mathbf{x}^* = \arg \min_{\mathbf{x}} \sum_{i,j} \left( \mathbf{x}^T \mathbf{B}_T^T \mathbf{C}_i^T \mathbf{C}_j \mathbf{B}_T \mathbf{x} - \delta_{ij} \right)^2,
\]

where \( \mathbf{C}_* \) is column selecting matrix defined in (6).

4. Equivalent inputs/outputs are

\[
\left( \begin{array}{c} \mathbf{u}_1^* \\ \mathbf{y}_1^* \end{array} \right) = \mathbf{B}_u \mathbf{x}^*.
\]

**C. Models Merging**

The advantage of equivalent data is that they are time consistent sequence allowing to build a new regressor and a RHS vector of any structure.

Assume merging of \( k \) ARX (FIR) models. Parameters mean value \( \mathbf{\hat{\theta}}_i \) and normalized covariances \( \mathbf{P}_i \) of each model are used to generate equivalent data of a selected length \( N \). These data are reordered to appropriate regressors \( \mathbf{Z}_i \) and RHS vectors \( \mathbf{b}_i \) of the same size according to selected target structural parameters \( n_a, n_b \) and stacked together

\[
\mathbf{Z} = \left( \begin{array}{c} \mathbf{Z}_1^T \\ \vdots \\ \mathbf{Z}_k^T \end{array} \right)^T, \quad \mathbf{b} = \left( \begin{array}{c} \mathbf{b}_1^T \\ \vdots \\ \mathbf{b}_k^T \end{array} \right)^T.
\]

Estimating merged model parameters by Maximum Likelihood leads to standard weighted least squares as the models have different residual variances \( s_i^2 \). Each block is weighted
by \( w_i = 1/s_i^2 \)

\[
\hat{\theta} = (Z^T W Z)^{1/2} Z^T W b,
\]

\[
P = (Z^T W Z)^{1/2},
\]

\[
s^2 = \frac{1}{\text{tr}(W)} \left( b - Z \hat{\theta} \right)^T W \left( b - Z \hat{\theta} \right),
\]

\[
W = \text{diag}(w_1, \ldots, w_k) \otimes I_N.
\]

Algorithm 2 ARX (FIR) models merging

**Inputs:**

- \( \theta_i, P_i \) means and covariances, \( i = 1, \ldots, k \)
- \( n_o, n_b, \) target structural parameters

**Outputs:**

- \( \hat{\theta}, P \) merged model mean and covariance

1. Select a length of equivalent data \( N \)
2. Generate equivalent data by Algorithm 1 for each model
   \[
   \hat{\theta}_i, P_i \rightarrow \left\{ u^{(i)}, y^{(i)} \right\}_1^N
   \]
3. Create regressors \( Z_i \) and RHS vectors \( b_i \) according to target structural parameters and stack them together (7).
4. Get merged model parameters by Maximum Likelihood leading to weighted least squares (8).

IV. MODELS Merging example

The example in Figure 3 shows merging of three models with different orders. The individual models are identified by PEM with ARX structure and data are generated by 4th order ARX system

\[
\begin{align*}
B(s) &= \frac{\omega_1^2 \omega_2^2}{s^2 + 2 \omega_1 \xi_1 s + \omega_1^2} \left( s^2 + 2 \omega_2 \xi_2 s + \omega_2^2 \right), \\
A(s) &= \left( s^2 + 2 \omega_1 \xi_1 s + \omega_1^2 \right) \left( s^2 + 2 \omega_2 \xi_2 s + \omega_2^2 \right),
\end{align*}
\]

with \( \omega_1 = 2, \omega_2 = 5, \xi_1 = 0.3, \xi_2 = 0.1 \), discretized with \( T_s = 0.2 \) s. Identification data for each model are obtained on the different part of frequency spectra (Figure 4) with signal to noise ratio 30 dB. The length of identification data is 300 and the length of equivalent data (Figure 5) was selected as 50. The properties of each model are summarized in the following table.

<table>
<thead>
<tr>
<th>Order</th>
<th>Inputs Excitation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>2 around 1st resonance frequency</td>
</tr>
<tr>
<td>Model 2</td>
<td>3 around 2nd resonance frequency</td>
</tr>
<tr>
<td>Model 3</td>
<td>1 low frequencies</td>
</tr>
</tbody>
</table>

A. Estimating Frequency Uncertainty

Frequency uncertainty in Figure 3 is obtained from second order Taylor expansion of \( G(\omega, \theta) \) around \( \hat{\theta} \)

\[
E \{ G(\omega) \} = \left. G(\omega, \hat{\theta}) + \frac{1}{2} \text{tr}(H(\omega) P) \right|_{\omega = \hat{\theta}},
\]

\[
\text{var} \{ G(\omega) \} = \left. g(\omega)^T P g(\omega) + \frac{1}{2} \text{tr}(H(\omega) P H(\omega) P) \right|_{\omega = \hat{\theta}},
\]

where

\[
g(\omega) = \left. \frac{\partial G(z, \theta)}{\partial \theta} \right|_{z = e^{j\omega T_s}, \theta = \hat{\theta}},
\]

\[
H(\omega) = \left. \frac{\partial^2 G(z, \theta)}{\partial \theta \partial \theta^T} \right|_{z = e^{j\omega T_s}, \theta = \hat{\theta}}.
\]
V. UNCERTAIN MODELS INTERCONNECTION

This section treats the problem of consistent parameters estimation of interconnected uncertain models. Single input single output ARX or FIR models with normally distributed parameters are assumed. Moreover parameters between models can be fully correlated as SISO models can be a result of MIMO models decomposition.

Assume n arbitrarily connected SISO systems (parallel, serial, feedback connections) with models

\[ M_i(d) = \frac{(d^{b_0} \ldots d^{b_{n_i}})}{(d^{a_0} \ldots d^{a_{n_i}})} \hat{b}_i, \]

where \( b_i \) and \( a_i \) are vectors of numerator and denominator polynomial coefficients. Models interconnection is represented by matrix \( C \) where \( C(i,j) = \pm 1 \) if \( M_j \) output enters \( M_i \) input. External input and output to interconnected models is determined by vectors \( m_{in} \) and \( m_{out} \) with "1" on proper positions. The transfer function of the whole model is

\[ M(d) = m_{out}^T (I - MC)^{-1} M_{in}, \]

where \( M(i,i) = M_i \) is diagonal. Resulting \( M(d) \) is rational function

\[ M(d) = \frac{(d^{b_0} \ldots d^{b_n})}{(d^{a_0} \ldots d^{a_n})} \hat{b} \]

with coefficient vectors \( b \) and \( a \) given as sum of convolutions

\[ b = \sum a \hat{i}_1 
\]

\[ a = \sum a \hat{j}_1 \]

where \( \hat{i}_1 \) and \( \hat{j}_1 \) are 0/1 selection variables. Statistically correct mean value and covariance of vectors \( b \) and \( a \) needs an ability to compute a mean value and a covariance of multiple normal vectors convolution.

A. Normal vectors convolution

Assume two random vectors \( h_1 \in \mathbb{R}^{n_1}, \ h_2 \in \mathbb{R}^{n_2} \) described by the first two moments

\[ h = \left( \begin{array}{c} h_1 \\ h_2 \end{array} \right), \quad \mathbb{E}\{h\} = \mu, \quad \text{var}\{h\} = \Sigma. \]

A convolution \( h_s = h_1 \ast h_2 \) can be written as a quadratic form

\[ h_s(i) = h^T Q_i h, \quad Q_i = \begin{pmatrix} 0_{n_1} & \bar{Q}_i \\ \bar{Q}_i^T & 0_{n_2} \end{pmatrix}, \]

for \( i = 1, \ldots, n_1 + n_2 - 1 \) and where

\[ \bar{Q}_i(p,q) = \begin{cases} 1/2 & \text{if } p + q = i + 1, \\ 0 & \text{otherwise}. \end{cases} \]

Using the relations for expectations of quadratic and quartic forms [10] the first two moments of convolution are
\[ \mu_* (i) = E \{ h_* (i) \} = E \{ h^T Q_i h \} \]
\[ = \text{tr}(Q_i \Sigma) + \mu^T Q_i \mu, \]
\[ \Sigma_{**} (i, j) = \text{cov} (h_* (i), h_* (j)) \]
\[ = E \{ h^T Q_i h h^T Q_j h \} - E \{ h^T Q_i h \} E \{ h^T Q_j h \} \]
\[ = 2 \text{tr}(Q_i \Sigma Q_j) + 4 \mu^T Q_i \Sigma Q_j \mu, \]
\[ \Sigma_* (i, j) = \text{cov} (h(i), h_* (j)) \]
\[ = E \{ s_i^T h h^T Q_j h \} - E \{ s_i^T h \} E \{ h^T Q_j h \} \]
\[ = 2 s_i^T \Sigma Q_j \mu, \]

where
\[ s_i (k) = \begin{cases} 1 & \text{if } k = i, \\ 0 & \text{otherwise.} \end{cases} \]

In a compact form
\[ E \left( \begin{pmatrix} h \\ h_* \end{pmatrix} \right) = \begin{pmatrix} \mu \\ \mu_* \end{pmatrix}, \quad \text{var} \left( \begin{pmatrix} h \\ h_* \end{pmatrix} \right) = \begin{pmatrix} \Sigma & \Sigma_* \\ \Sigma^T & \Sigma_{**} \end{pmatrix}. \] (10)

Multiple vectors can be convoluted sequentially
\[ h_* = h_1 * \cdots * h_n = \left( (h_1 * h_2) * h_3 \right) * \cdots * h_n, \]
by using the previous results. Correct summing of convolution terms in (9) needs a knowledge of cross-covariances, which are easily obtained by extending vector \( h \) in (10) as
\[ h' = \begin{pmatrix} h \\ h_* \end{pmatrix}, \]

after computing each convolution. This preserves cross-covariances between all terms and allows to compute final statistics of numerator and denominator vectors by using formulas for summing correlated variables [10]. Finally it has to be noted that Gaussian or Gauss-inverse-Wishart distribution of the original random vectors is lost in the convolution.

**B. Example**

Low complexity example with a connection of five uncertain models is in Figure 6. The models \( M_1, M_3, M_5 \) are of the 2nd order and the models \( M_2, M_4 \) are of the 1st order with mutually correlated normal coefficients. The level of uncertainty is exaggerated to demonstrate the difference between statistically correct computations (denoted as proper connection) and computations with mean values only (denoted as simple connection). A comparison of the results obtained by ignoring models uncertainty and by proper models combination is in Figure 7.

![Fig. 6. Models interconnection](image)

VI. CONCLUSION

The Presented algorithms were developed as a part of platform for modeling and identification of large-scale systems. Multiple sub-models are consistently combined into the global model and overlapping information in sub-models require tool for consistent merging of multiple models of a single system.

Models merging can be also advantageous for data sets, where conditions (noise characteristics) changed in time or in a case of multiple data sets obtained under different conditions. This contradicts usual assumption of constant noise characteristics. Dividing data, identifying multiple models and merging them together gives better results as noise levels for each model are estimated separately.

Another use is an on-demand model actualization, where an actual model is merged with a model obtained by exciting system inputs on the frequency spectrum where the model predictions quality was detected as low.

VII. ACKNOWLEDGMENT

The authors gratefully acknowledge the financial support of European Commission FP7 project "WIDE" (project no. 224168).

REFERENCES


![Fig. 7. Comparison of combined model poles computed by the different methods.](image)
