Integrating Prior Information into Subspace Identification Methods

Pavel Trnka and Vladimír Havlena

***Abstract*— Integrating prior information into subspace identification methods improves their usability for industrial data, where experimental data by them self are in many cases not good enough to give a proper model. The identification experiments in the industrial environment are limited by the economical and safety reasons. However, in practical applications, there is often strong prior information about the identified system, which can be exploited in the identification. The presented algorithm formulates subspace identification as a multi-step predictor optimization. Reformulation to the Bayesian framework allows to incorporate prior information. The paper is completed with the application to the experimental data from the oil burning steam boiler with the rated power of 100 MW.**

I. INTRODUCTION

The good properties of Subspace State Space System IDentification (4SID) have shown their usability in the industrial applications [1]. Mainly their robustness and ability to identify MIMO (Multiple Inputs Multiple Outputs) systems with the same complexity as for SISO (Single Input Single Output) systems without the need for extensive structural parametrization (as for example for MIMO ARX models).

However, the experimental input/output data by them self are in many cases not good enough to give a proper model. This may be caused by the fact, that the identification experiments in the industrial environment are limited by the economical and safety reasons, which results into I/O data without proper excitation and with strong noise burden. The black-box identification approach, such as in 4SID, relying only on the measured data, may fail in such cases.

However, in the practical applications there is often strong prior information about the system, which should be exploited by the identification algorithm to significantly improve the identified model quality. Such information should be: an approximate knowledge of time constants, the known static gains, an integrating character etc. Incorporating prior information into 4SID methods will be addressed in this article.

The presented algorithm uses the formulation of subspace identification algorithm as a multi-step predictor optimization. The non-causality and over-parametrization are eliminated by a convenient problem reformulation [2]. The prior information incorporation is allowed by putting 4SID into Bayesian framework.

The paper is organized as follows. First the notation is established, next the unified subspace identification algorithm is shown and the optimality criterion is briefly derived. Next, the algorithm incorporating prior information is proposed and different types of useful prior information are shown. The paper is concluded with practical application of the algorithm.

This work was supported by financial support of the Grant Agency of Czech Republic under grant No. 102/05/0271 and 102/05/2075.

P. Trnka is with the Department of Control Engineering, Czech Technical University in Prague, Karlovo náměstí 13, 121 35 Prague 2, Czech Republic trnkap@control.felk.cvut.cz

V. Havlena is with Honeywell Prague Laboratory, Honeywell Intl., Pod vodárenskou věží 4, 182 08 Prague 8, Czech Republic havlena@htc.honeywell.cz

II. NOTATION AND OVERVIEW

The first idea in 4SID methods, is to combine a state space model and a set of I/O data into *a single linear matrix equation, relating the signal matrices with the parameters matrices*. Prior to this, some definitions are necessary.

A. State Space Model

In this paper a state space model of stochastic system in the innovation form [3] is considered

$$x_{k+1} = Ax_k + Bu_k + Ke_k, \quad (1)$$

$$y_k = Cx_k + Du_k + e_k, \quad (2)$$

where $u_k \in \mathbb{R}^m$ is the m-dimensional input, $x_k \in \mathbb{R}^n$ is the n-dimensional state, $y_k \in \mathbb{R}^l$ is the l-dimensional output, K is the steady state Kalman gain and $e_k \in \mathbb{R}^l$ is an unknown innovation with $E[e_k] = 0$ and covariance matrix $E\left[e_k e_k^T\right] = R_e$.

B. Signal Related Matrices

In 4SID algorithms, all signals (the inputs, the outputs and the noises) are arranged into Hankel matrices. Assume a known set of input/output data samples u_k,y_k for $k \in \langle 0, 1, \ldots, i+h+j-2\rangle$. These samples can be arranged into block Hankel matrices with i and h block rows and j columns as follows

$$\left(\frac{U_p}{U_f}\right) = \left(\begin{array}{cccc} u_0 & u_1 & \ldots & u_{j-1} \\ u_1 & u_2 & \ldots & u_j \\ \vdots & \vdots & \ddots & \vdots \\ u_{i-1} & u_i & \ldots & u_{i+j-2} \\ \hline u_i & u_{i+1} & \ldots & u_{i+j-1} \\ u_{i+1} & u_{i+2} & \ldots & u_{i+j} \\ \vdots & \vdots & \ddots & \vdots \\ u_{i+h-1} & u_{i+h} & \ldots & u_{i+h+j-2} \end{array}\right)$$

where $U_p \in \mathbb{R}^{im\times j}$ is the matrix of past inputs and $U_f \in \mathbb{R}^{hm\times j}$ is the matrix of future inputs. Although most data samples can be found in both matrices, the corresponding columns of U_p and U_f are subsequent without any common

data samples. The distinction between the past and the future is important for the predictor and Kalman filter concept used in 4SID.

The values of the coefficients i and h are usually selected slightly larger than the upper bound of expected system order and the coefficient $j = s - i - h + 1$, where s is the number of available data samples.

For the outputs y_k and the noises e_k similar Hankel matrices Y_p, Y_f and E_p, E_f can be constructed. A combination of U_p and Y_p denoted as W_p is used as a regressor

$$W_p = \begin{pmatrix} Y_p \\ U_p \end{pmatrix}$$

The system state sequence is also used in a matrix form with the following structure

$$X_p = (x_0 \; x_1 \; \ldots \; x_{i-1}), \quad X_f = (x_i \; x_{i+1} \; \ldots \; x_{i+j-1}).$$

C. Parameters Related Matrices

The extended observability matrix Γ_k is an extension of observability matrix for a number of block rows higher than the system order $k \geq n$

$$\Gamma_k = \begin{pmatrix} C \\ CA \\ \vdots \\ CA^{k-1} \end{pmatrix} \in \mathbb{R}^{kl \times n}.$$

Similarly *the reverse extended controllability matrix* Δ_k^d for the deterministic input

$$\Delta_k^d = \begin{pmatrix} A^{k-1}B & A^{k-2}B & \ldots & B \end{pmatrix} \in \mathbb{R}^{n \times km}.$$

The block Toeplitz matrix H_k^d composed from the impulse responses elements $\{g_0, \ldots, g_{k-1}\}$

$$H_k^d = \begin{pmatrix} g_0 & 0 & \ldots & 0 \\ g_1 & g_0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ g_{k-1} & g_{k-2} & \cdots & g_0 \end{pmatrix} \in \mathbb{R}^{kl \times km}. \quad (3)$$

D. Single Equation Formulation of State Space Model

As already mentioned before, the starting point of 4SID methods is a reformulation of the recursive state space innovation model (1,2) and a set of experimental data into one single matrix equation, sometimes denoted as extended state space model. This is done by recursive substitution of (1) into (2)

$$Y_p = \Gamma_i X_p + H_i^d U_p + H_i^s E_p, \quad (4)$$
$$Y_f = \Gamma_h X_f + H_h^d U_f + H_h^s E_f, \quad (5)$$
$$X_f = A^i X_p + \Delta_i^d U_p + \Delta_i^s E_p. \quad (6)$$

Equations (4) and (5) are similarly defining outputs as a linear combination of previous states by the extended observability matrix $\Gamma_\bullet$ (response from the states) and a linear combination of previous inputs and noises by their respective impulse responses $H_\bullet^d$ and $H_\bullet^s$. Equation (6) is relating the future and the past states under the influence of the inputs and the noises.

III. 4SID IDENTIFICATION

This section recalls the basic unified 4SID algorithm proposed by Overschee and De Moore [4].

A. Problem formulation

The solved problem can be formulated as follows:

Given s samples of the input sequence $\{u_0, \ldots, u_{s-1}\}$ and the output sequence $\{y_0, \ldots, y_{s-1}\}$.

Estimate the parameters of the state space model in the innovation form (1,2), i.e. estimate the system order n and obtain the parameters A, B, C, D, K up to within a similarity transformation and covariance matrix R_e of the noise e_k.

B. Unified 4SID algorithm

First step is the computation of the oblique projection [5]. The row space of future outputs Y_f is projected on the row space of past data W_p along the row space of future inputs U_f

$$\mathcal{O}_h = Y_f \underset{U_f}{/} W_p. \quad (7)$$

Having obtained the matrix $\mathcal{O}_h$, the rest of the algorithm uses the fact, that $\mathcal{O}_h$ can be factorized as

$$\mathcal{O}_h = \Gamma_h \hat{X}_f,$$

Exploiting this fact the weighted matrix $\mathcal{O}_h$ is factorized by the singular value decomposition (SVD)

$$W_1 \mathcal{O}_h W_2 = U \Sigma V^T.$$

The weights W_1 and W_2 allow for tuning the algorithm, i.e. selecting a particular method (N4SID, MOESP, CVA). The order n of the system is determined by inspecting the singular values in Σ and used to partition U, Σ and V^T to $U_1 = U(:, 1:n)$, $\Sigma_1 = \Sigma(1:n, 1:n)$ and $V_1^T = V(:, 1:n)^T$ (Matlab like notation). Then

$$\begin{aligned} \Gamma_h &= W_1^{-1} U_1 \Sigma_1^{1/2}, \\ \hat{X}_f &= \Gamma_h^\dagger \mathcal{O}_h. \end{aligned}$$

From the knowledge of the estimated state sequence $\hat{X}_f$ and input/output data, the state space model parameters A, B, C and D can be computed by the least squares or total least squares from

$$\begin{pmatrix} \hat{X}_{i+1} \\ Y_i \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} \hat{X}_i \\ U_i \end{pmatrix} + \varepsilon,$$

where Y_i is first block row of Y_f and similarly U_i. Finally the stochastic properties can be estimated from the residuals

$$\begin{aligned} \hat{R}_e &= \Sigma_{22}, \\ \hat{K} &= \Sigma_{12} \Sigma_{22}^{-1}, \end{aligned} \quad \text{where} \quad \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} = \operatorname{cov}(\varepsilon).$$

IV. 4SID AS MULTI-STEP OPTIMAL PREDICTOR

This section recalls, that optimality criterion of 4SID is a *minimization of multi-step predictions error based on input/output data*. And that the optimal solution leads to the oblique projection used as a corner stone of unified 4SID algorithm [6].

A. Multi-step predictor

Assume the state space model is known and at time i, a system state x_i and a sequence of future inputs $\{u_i, u_{i+1}, \ldots, u_{i+h-1}\}$ are also known. The output predictions for 0 to $h-1$ steps ahead can be estimated as (the innovations are replaced by their mean value $E[e_k] = 0$)

$$\begin{pmatrix} \hat{y}_i \\ \vdots \\ \hat{y}_{i+h-1} \end{pmatrix} = \Gamma_h x_i + H_h^d \begin{pmatrix} u_i \\ \vdots \\ u_{i+h-1} \end{pmatrix}.$$

The output predictions for j subsequent initial states can be written in a compact form using Hankel matrices

$$\hat{Y}_f = \Gamma_h X_f + H_h^d U_f, \tag{8}$$

where every column of $\hat{Y}_f$ represents a sequence of linear output predictions based on the initial state and a sequence of inputs from the corresponding columns of X_f and U_f. The states X_f, unknown in the process of identification, can be estimated from the past data W_p as [6]

$$\hat{X}_f = \begin{pmatrix} A^h \Gamma_i^\dagger & (\Delta_h^d - A^h \Gamma_i^\dagger H_i^d) \end{pmatrix} \begin{pmatrix} Y_p \\ U_p \end{pmatrix} = L_w' W_p.$$

Using the estimated states, the predictor (9) becomes

$$\widetilde{Y}_f = L_w W_p + H_h^d U_f, \tag{9}$$

where $\widetilde{Y}_f$ is the output linear estimate of Y_f using *a finite available input/output data set.*

B. Multi-step predictions optimization

Consider the optimal multi-step predictions on a given set of input/output data. That is to find the parameter matrices L_w and H_h^d of (9) to optimally predict the outputs. The quality of the predictions will be measured by a Frobenius norm of prediction errors

$$\min_{L_w, H_h^d} \left\| Y_f - \widetilde{Y}_f \right\|_F = \min_{L_w, H_h^d} \left\| Y_f - \begin{pmatrix} L_w & H_h^d \end{pmatrix} \begin{pmatrix} W_p \\ U_f \end{pmatrix} \right\|_F. \tag{10}$$

Denoting $\mathcal{D} = \begin{pmatrix} W_p^T & U_f^T \end{pmatrix}^T$ and assuming a persistent excitation, the optimal L_w and H_h^d can be obtained as

$$\begin{pmatrix} L_w & H_h^d \end{pmatrix} = Y_f \mathcal{D}^T \left(\mathcal{D} \mathcal{D}^T \right)^{-1}.$$

Using the previous result, the estimated zero-input state response $L_w W_p$ can be written as

$$L_w W_p = Y_f \mathcal{D}^T \left[\left(\mathcal{D} \mathcal{D}^T \right)^{-1} \right] \begin{pmatrix} \mathbf{I}_{r \times r} \\ \mathbf{0}_{hm \times r} \end{pmatrix} W_p, \quad r = i(l+m),$$

which is the expression for the oblique projection

$$L_w W_p = Y_f \underset{U_f}{/} W_p,$$

showing the equivalency between the oblique projection in 4SID and an optimal multi-step predictor parameters estimation.

$$\mathcal{O}_i = \Gamma_i \hat{X}_f = Y_f \underset{U_f}{/} W_p = L_w W_p.$$

C. Enforcing causality and parameters uniqueness

The parameter matrix H_h^d obtained from the oblique projection (7) may not have the block Toeplitz structure with zeros above the main diagonal according to (3), which leads to predictor non-causality. A solution was proposed in [2]. To enforce the causality and the parameters uniqueness it applies a formula for the vectorization of matrix product to (9)

$$\operatorname{vec}(Y_f) = \left(\begin{pmatrix} W_p^T & U_f^T \end{pmatrix} \otimes I \right) \operatorname{vec}\left(\begin{pmatrix} L_w & H_h^d \end{pmatrix} \right)$$

and uses the fact, that it is possible to find N that

$$\operatorname{vec}\left(\begin{pmatrix} L_w & H_h^d \end{pmatrix} \right) = N \begin{pmatrix} \ell_w \\ g \end{pmatrix},$$

where

$$\ell_w = \operatorname{vec}(L_w), \quad g = \operatorname{vec}\left(\begin{pmatrix} g_0 & \ldots & g_{h-1} \end{pmatrix} \right).$$

A set of equations equivalent to (9) with enforced H_h^d structure is then

$$\underbrace{\operatorname{vec}(Y_f)}_{y} = \underbrace{\left(\begin{pmatrix} W_p^T & U_f^T \end{pmatrix} \otimes I \right) N}_{Z} \underbrace{\begin{pmatrix} \ell_w \\ g \end{pmatrix}}_{\theta}. \tag{11}$$

V. Incorporating prior informaion

A natural tool for working with prior information is Bayesian framework [7]. It allows to combine prior information with information from the experimental data.

In order to use Bayesian inference, there have to be the prior probability density function (p.d.f.) of parameters θ and the likelihood function $l(\theta|y)$ of parameters conditioned by the experimental data y.

The prior information has to be transformed into the prior p.d.f. of parameters. Assuming a normal distribution it will be described as

$$p_{\text{prior}}(\theta) = N(\hat{\theta}_0, P_0)$$

and real-world prior information will be transformed into $\hat{\theta}_0$ and P_0 by a covariance matrix shaping.

The likelihood function $l(\theta|y)$ is determined by the rearranged multi-step predictor model (11), with a simple structure

$$Z\theta = y + e, \qquad e \sim \mathrm{N}(0, R).$$

The posterior p.d.f of θ can be found from Bayesian rule

$$p_{\text{post}}(\theta) \propto l(\theta|y) p_{\text{prior}}(\theta).$$

For our chosen $p_{\text{prior}}(\theta)$ and $l(\theta|y)$ it leads to the linear mean square error estimate (LMS)

$$\begin{aligned} \hat{\theta} &= \theta_0 + P_0 Z^T \left(Z P_0 Z^T + R \right)^{-1} (y - Z\theta_0), \\ P_\theta &= P_0 - P_0 Z^T \left(Z P_o Z^T + R \right)^{-1} Z P_0. \end{aligned}$$

A. Covariance matrix shaping

To describe prior information by $\hat{\theta}_0$ and P_0, there is an instrument usually denoted as a covariance matrix shaping [8]. The idea is to reduce the uncertainty of the parameters in the directions with well known prior information.

Assume that there is strong prior information on the parameters in the direction r. The corresponding covariance matrix P should have small eigenvalue $\sigma_1 \ll 1$ in r direction and large other eigenvalues $\sigma_0 \gg 1$

$$P = \sigma_1 \frac{rr^T}{r^T r} + \sigma_0 \left(I - \frac{rr^T}{r^T r} \right).$$

The following sections show some examples of r determination for some different types of prior information. The covariance matrices P_i describing different prior information are combined to one covariance P as

$$P^{-1} = \sum_i P_i^{-1}.$$

B. Known static gain

Assume a known static gain K from the p-th input to the q-th output $u(p) \to y(q)$

$$K = f(\theta) = \sum_{i=0}^{h-1} g_i(q,p) = \text{const.} \tag{12}$$

This affine parameters constraint can be represented by any θ_0 satisfying (12) such as

$$\theta_0 = \left(\begin{array}{c|c} \ell_w^T & g^T \end{array} \right)^T = \frac{K}{h} \left(\begin{array}{c|c} \mathbf{0} & \underbrace{w^T \quad \dots \quad w^T}_{\text{h times}} \end{array} \right)^T, \tag{13}$$

where $w \in \mathrm{R}^{ml}$

$$w(k) = \begin{cases} 1 & \dots \quad k = p + (q-1)m, \\ 0 & \dots \quad \text{otherwise} \end{cases}$$

and P_0 with small eigenvalue in the direction of $f(\theta)$ gradient

$$r = \nabla f(\theta) = \left(\begin{array}{c|c} \mathbf{0} & \underbrace{w^T \quad \dots \quad w^T}_{\text{h times}} \end{array} \right)^T, \tag{14}$$

and large other eigenvalues. This ensures, that the experimental data will change the parameter values freely in the directions perpendicular to r and will have only small influence to the static gain (depending on the ratio of σ_0/σ_1).

The concept can be used to any prior information describable as an affine combination of parameters

$$f(\theta) = \sum k(i)\theta(i) = c$$

or generalized to a nonlinear function for a good prior estimate of θ_0.

C. Known ratio between static gains

Assume that the particular gain values for two inputs ($u(p_1) \to y(q)$ and $u(p_2) \to y(q)$) are unknown, but their relative ratio is well known (arises in practical applications)

$$K = f(\theta) = \frac{\sum_{i=0}^{h-1} g_i(q,p_1)}{\sum_{i=0}^{h-1} g_i(q,p_2)} = \text{const.} \tag{15}$$

This constraint can be represented by θ_0 given by (13) satisfying (15) with w as

$$w(k) = \begin{cases} K & \dots \quad k = p_1 + (q-1)m, \\ 1 & \dots \quad k = p_2 + (q-1)m, \\ 0 & \dots \quad \text{otherwise} \end{cases}$$

and P_0 with small eigenvalue in the direction of $f(\theta)$ gradient (14), which for the chosen values of θ_0 simplifies to

$$r = \nabla f(\theta) = \left(\begin{array}{c|c} \mathbf{0}_{1\times 2hi} & \underbrace{v^T \quad \dots \quad v^T}_{\text{h times}} \end{array} \right)^T,$$

where

$$v(k) = \begin{cases} 1 & \dots \quad k = p_1 + (q-1)m, \\ -K & \dots \quad k = p_2 + (q-1)m, \\ 0 & \dots \quad \text{otherwise.} \end{cases}$$

D. Ensuring smoothness of step response

One of the natural request to the identified impulse or step response is the smoothness. It means to suppress the fitting of high-frequency disturbances. The smoothness will be shown for SISO system. The idea is that the smooth impulse response has small second order differences

$$\Delta^2 g_i = g_{i+2} - 2g_{i+1} + g_i \to 0.$$

The differences can be written as

$$\begin{pmatrix} 1 & & & & \\ -2 & 1 & & & \\ 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1 \end{pmatrix} g = Dg \to 0.$$

Covariance matrix P_0 should be chosen so that $\operatorname{cov}\{Dg\}$ is sufficiently small

$$\operatorname{cov}\{Dg\} = DP_gD^T \approx \sigma_h^2 I, \quad \sigma_h \to 0,$$

giving P_0 as

$$P_0 = \begin{pmatrix} \sigma_0 \mathrm{I}_{2hi\times 2hi} & 0 \\ 0 & \sigma_h^2 D^{-1} D^{-T} \end{pmatrix},$$

where σ_h^2 is the tuning parameter enforcing the smoothness for small values. The smooth initial estimate of θ_0 can be obtained from the impulse response of the first order model or simply as $\theta_0 = 0$.

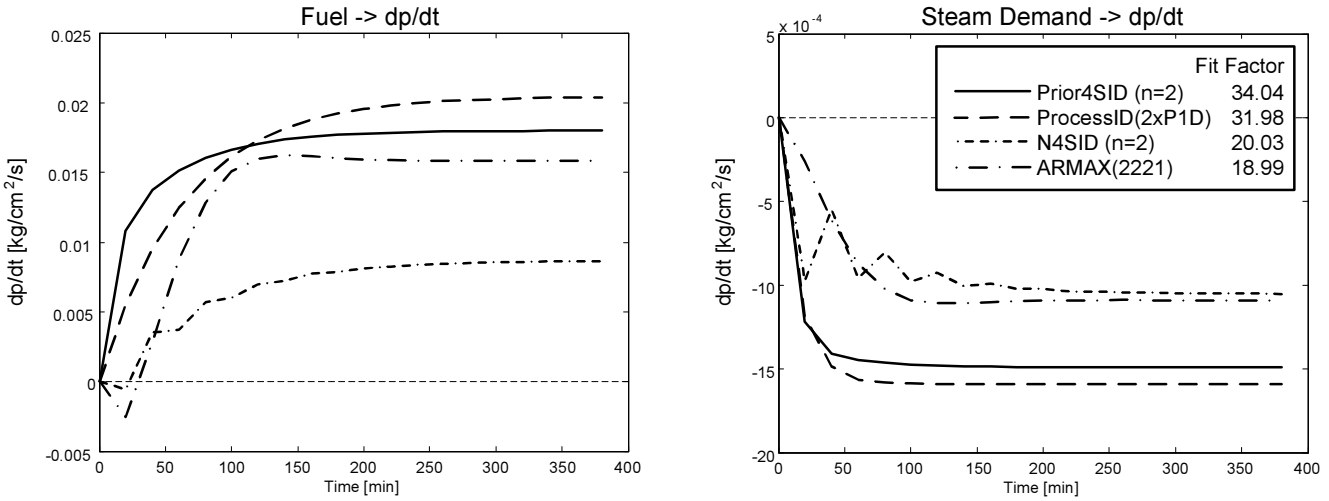

Fig. 1. Step responses of the models identified by 4SID with prior information (n=2), Process Identification (numerical PEM on 2xP1D structure), N4SID (n=2) and PEM on ARMAX model (2221).

VI. State space model realization with prior information

Estimating the state space model parameters from the impulse response sequence is traditionally done by classical Ho & Kalman realization theory [9]. However, LMS gives an estimate of the impulse response $\hat{G} = \{\hat{g}_0, \ldots, \hat{g}_{h-1}\}$ and its *covariance matrix* P_G (containing the non-diagonal values). The information contained in the covariance matrix is important for a proper state realization and must be used in the realization algorithm in order to *preserve the supplied prior information* and to respect the experimental data quality.

First consider the non-diagonal values of P_G to be zero. The solution is similar to Ho and Kalman realization. Construct a Hankel matrix from the impulse responses

$$T = \begin{pmatrix} \hat{g}_1 & \hat{g}_2 & \cdots & \hat{g}_p \\ \hat{g}_2 & \hat{g}_3 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \hat{g}_p & \cdots & \cdots & \hat{g}_{h-1} \end{pmatrix}.$$

Then instead of usual lower rank approximation by SVD, the Weighted Lower Rank-n Approximation (WLRA) has to be used

$$T' = \arg\min_{\substack{T' \\ \text{rank}(T')=n}} \|W. * (T - T')\|_F, \qquad (16)$$

where $.*$ is an element-wise multiplication and W are the weights for each impulse response element from P_G

$$W = \begin{pmatrix} P_G(1,1)^{-1} & P_G(2,2)^{-1} & \cdots & P_G(p,p)^{-1} \\ P_G(2,2)^{-1} & P_G(3,3)^{-1} & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ P_G(p,p)^{-1} & \cdots & \cdots & P_G(h\text{–}1,h\text{–}1)^{-1} \end{pmatrix}.$$

The (16) can be solved by the algorithm described in [10]. Obtaining T' as a weighted rank-n approximation of T, the rest of Ho & Kalman realization algorithm can be used to get the state space model parameters A',B',C',D'. Using Matlab like notation

$$\begin{aligned} T' &= U\Sigma V^T, \\ \Gamma &= U(:, 1:n)\Sigma(1:n, 1:n)^{1/2}, \\ \Delta &= \Sigma(1:n, 1:n)^{1/2} V(1:n, :)^T, \\ A' &= \Gamma(1:end-l, :) \backslash \Gamma(l+1:end, :), \\ B' &= \Delta(1:m, 1:n), \quad C' = \Gamma(1:l, 1:n), \quad D' = \hat{g}_0. \end{aligned}$$

To incorporate the non-diagonal covariance elements the realization problem can be formulated in the sense of constrained maximum likelihood as

$$\min_{A,B,C,D} \left(\hat{G} - G\right)^T P_G^{-1} \left(\hat{G} - G\right), \qquad (17)$$

where

$$G = \begin{pmatrix} D & CB & CAB & \ldots & CA^{N-2}B \end{pmatrix}.$$

This optimization is nonlinear and the numerical solution is likely to converge to the local extreme. Practical experience is that when started from the solution A',B',C',D' given by scalar WLRA, the global extreme is almost always reached and the state space model parameters A,B,C,D are found.

VII. Algorithm brief summary

The following steps briefly summarize the steps of proposed algorithm.

1) Transform available prior information to the initial estimates of θ_0 and P_0. Several types of prior information can be mixed together.
2) Compute the posterior parameter estimates $\hat{\theta}$, P_θ from (12).
3) Using WLRA and Ho & Kalman realization find the first model parameter estimate A', B', C', D'.
4) Re-optimize A', B', C', D' by (17) to get the state space model parameters A, B, C, D.

VIII. Practical application

The algorithm was applied to the experimental data from the oil firing steam boiler with the rated effective power of 100 MW. The goal was to identify a model relating the fuel flow and the steam demand to the boiler pressure. This is known identification problem [11]. Both inputs are typically strongly collinear, because technological limitations do not allow independent excitation of each input. Several prior information were used:

- Known relative gain between fuel flow $\rightarrow$ drum pressure and steam demand $\rightarrow$ drum pressure. The value was obtained from the steady state boiler operation data by total least squares (Figure 2).
- Strictly dynamic character $g(0) = (0 \;\; 0)$.
- Step responses smoothness.

The model identified by 4SID with prior information gives the best model fit and is the only one to correspond with known prior information (Figure 1).

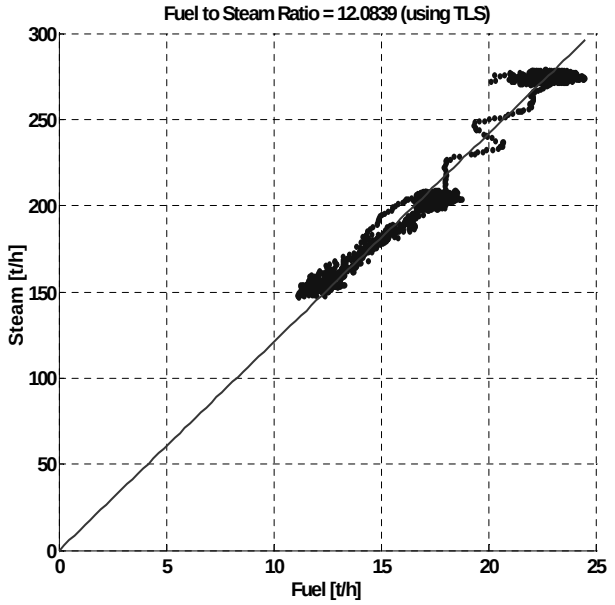

Fig. 2. Estimating the fuel flow to steam flow ratio by total least squares.

IX. Conclusion

The proposed algorithm incorporates prior information into subspace methods. It helps to improve the identification results in the practical applications, where prior information is often available and the experimental data may not be optimal. Moreover the algorithm allows for the recursification, because the parameters θ have fixed size and new measurements only append rows to the regressor Z.

X. Acknowledgment

The authors gratefully acknowledge financial support of the Grant Agency of Czech Republic under grant No. 102/05/0271 and 102/05/2075.

References

[1] W. Favoreel, B. D. Moor, and P. V. Overschee, "Subspace state space system identification for industrial processes," *Journal of Process Control*, no. 10, pp. 149–155, 2000.

[2] K. Peternell, W. Cherrer, and M. Deistler, "Statistical analysis of novel subspace identifiaction methods," *Signal Processing*, no. 52, pp. 161–177, February 1996.

[3] L. Ljung, *System Identification - Theory for the user*. Prentice-Hall, 1987, ISBN 0138816409.

[4] P. V. Overschee and B. D. Moor, "A unifying theorem for three subspace system identification," *Automatica, Special Issue on Trends in System Identification*, vol. 31, no. 12, pp. 1853–1864, 1995.

[5] D. A. Harville, *Matrix Algebra From a Statistician's Perspective*. Springer, 1997, ISBN 038794978X.

[6] P. V. Overschee and B. D. Moor, *Subspace Identification for Linear Systems: Theory-Implementation-Applications*. Kluwer Academic Publishers, 1996, ISBN 0792397177.

[7] V. Peterka, "Bayesian approach to system identification," in *Trends and Progress in System Identification*, P. Eykhoff, Ed. Pergamon Press, 1981, pp. 239–304.

[8] M. Kárný, "A note on feeding uncertain knowledge into recursive least squares," in *Proceedings of the 30th IEEE Conference on Decision and Control*, 1991, pp. 975–981.

[9] B. L. Ho and R. E. Kalman, "Effective construction of linear state variable models for input output functions," *Regelungstechnik*, no. 14, pp. 545–548, 1966.

[10] N. Srebro and T. Jaakkola, "Weighted low rank approximation," in *Proceedings of the 20th International Conference on Machine Learning*, 2003, pp. 720–727.

[11] K. J. Astrom and R. D. Bell, "Drum-boiler dynamics," *Automatica*, vol. 36, no. 3, pp. 363–378, 2000.