Proceedings of the International Conference of Control, Dynamic Systems, and Robotics
Ottawa, Ontario, Canada, May 15-16 2014
Paper No. 57

High-order Least Squares Identification

Rajamani Doraiswami
The University of New Brunswick
Fredericton, New Brunswick, Canada
dorai@unb.ca

Lahouari Cheded
Systems Eng. Department, King Fahd University of Petroleum & Minerals,
Dhahran 31261, KSA
cheded@kfupm.edu.sa

Abstract - In order to ensure that the estimates of system parameters are unbiased and efficient, most identification schemes including the Prediction Error Method (PEM), and the Subspace Method (SM), are based on minimizing the residual of the Kalman filter, and not the equation error (associated with system model) - as the residual is a zero mean white noise process whereas the equation error is coloured noise which may be correlated with data vector. The residual is linear in the input and the output of the system, and is nonlinear in the parameters to be estimated. The parameters enter in the expression for the residual as coefficients of rational polynomials associated with the input and the output. Similar to the PEM, which is a gold standard for comparing the performance identification schemes, the High Order Least Squares (HOLS) method is derived from the expression of the residual. In order to ensure that the equation error is a zero mean white noise process, the rational polynomials are approximated by finite high order polynomials by selecting the model order to be sufficiently high. As result, the relationship governing the residual and the parameters is linear, and the HOLS method becomes essentially a Least Squares (LS) method. A reduced order model is derived using frequency weighted LS approach. The performance of the HOLS is arbitrarily close to that of the PEM: estimates are unbiased and efficient. Unlike the PEM, the HOLS estimates as well the covariance of the estimation error have closed form expressions, that is, they are not computed iteratively. A reduced order model is derived using frequency weighted least squares approach. The proposed scheme has been successfully evaluated on a number of simulated and physical systems and favourably compared with the prediction error method (PEM).

***Keywords*:** High order least squares method, least squares method, prediction error method, subspace method, Kalman filter, residual, equation error.

1. Introduction

The LS method is widely used to identify a system as it is simple, numerically efficient, and yields a closed-form solution to the parameter estimate. If the equation error is a zero- mean white noise process, the estimate will be unbiased and efficient. However if the equation error is a colored noise, the estimate will be biased as the colored noise will be correlated with the data vector. To overcome this problem, approaches such as the PEM, the SM and the HOLS method have been proposed. The PEM is iterative and does not provide a closed-form solution to the parameter estimation problem.

A high-order model is used in various applications including the non-parameteric identification of impulse response, estimation of Markov parameters in the SM, in model predictive control, identification of a signal model and in system identification. The use of HOLS for identification of a signal model is inspired by the seminal paper by (Kumaresan & Tufts, 1982) for an accurate estimation of the parameters of an impulse response from measurements in an additive white noise. It is shown via simulation that the variance of the parameter estimation error approaches the Cramer-Rao lower bound. Further it is shown analytically that using a high-order model (with an order several times larger than the true order) improves significantly the accuracy of the parameter estimates. The HOLS has not received much attention in system identification although it has been mentioned as an alternative scheme to the PEM (Forssell & Ljung, 1999) ,(Ljung, 1999) and has been successfully employed in identification for fault diagnosis in (Doraiswami, Diduch, & Tang, 2010) , (Doraiswami & Cheded, 2013).

Most of the identification methods including the LS, PEM and SM use a Kalman filter model to derive the structure of the model set used in identification (Ljung, 1999). The residual of the Kalman filter takes over the combined role of the disturbance and the measurement noise affecting the system. As the Kalman filter residual, termed also an innovation process, is a zero-mean white process, one can develop an appropriate identification scheme using the vast literature on the statistical estimation theory for parameter estimation in a filtered white noise process (or colored noise), (Mendel, 1995).

The HOLS method similar to the PEM is developed starting with the derivation of the linear regression model based on the Kalman filter structure. The residual of the Kalman filter is expressed in terms of the numerator and the denominator polynomials of the system, and the *Kalman polynomial* (which is the denominator polynomial of the Kalman filter transfer function), (Doraiswami & Cheded, 2012). However, unlike in the case of PEM, in the HOLS method, the regression model is whitened by dividing the denominator and numerator polynomials by the Kalman polynomial. As the Kalman polynomial is stable, the coefficients of the polynomials resulting from the division operation are truncated to some finite, but large, number of terms. Although the equation error in the PEM is a zero-mean white noise process, it is not linear in the unknown parameters (formed of the coefficients of the numerator, denominator and Kalman polynomials). In the SM, a high- order model is derived from the Kalman filter state- space model by truncating the impulse response, which is expressed in terms of the Markov parameters, to some finite but large number of terms.

In many applications such as fault diagnosis, performance monitoring and controller design, a reduced- order model is desired. The commonly-used approach for obtaining a reduced-order model includes the balanced realization approach and the frequency- weighted LS estimator approach. The frequency-weighted LS estimator method is used to estimate of the true system model, termed here reduced-order model, which is derived from the high-order model by minimizing a frequency-weighted residual (Doraiswami, 2005).

2. Mathematical Model of the System and the Kalman Filter

The state-space model $\left(\boldsymbol{A},\boldsymbol{B},\boldsymbol{E}_w,\boldsymbol{C}\right)$, termed *process form*, of the system given by:

$$\begin{aligned} \boldsymbol{x}(k+1) &= \boldsymbol{A}\boldsymbol{x}(k)+\boldsymbol{B}r(k)+\boldsymbol{E}_w w(k) \\ y(k) &= \boldsymbol{C}\boldsymbol{x}(k)+v(k) \end{aligned} \tag{1}$$

where $\boldsymbol{x}(k)=\left[x_1(k) \quad . \quad x_n(k)\right]^T$ is an (nx1) vector of states, $r(k)$ the input , $w(k)$ the disturbance $v(k)$ the measurement noise , and $w(k)$ and $v(k)$ are zero-mean white noise processes ; $\boldsymbol{A}$ is (nxn) matrix, $\boldsymbol{B}$ an (nx1)vector, $\boldsymbol{E}_w$ an (nx1) disturbance entry vector and $\boldsymbol{C}$ a (1xn) vector; It is assumed that the system is controllable and observable. The difference equation model relating the input $r(k)$ and output $y(k)$ is given by:

$$y(k)=-\sum_{i=1}^{n} a_i y(k-i)+\sum_{i=1}^{n} b_i u(k-i)+\upsilon(k) \tag{2}$$

$$\upsilon(z)=\left(\boldsymbol{N}_w\left(z\right)w(z)+D(z)v(z)\right) \tag{3}$$

where, $\upsilon(k)$ is the equation error, $\boldsymbol{N}_w\left(z\right)=\boldsymbol{C}adj\left(z\boldsymbol{I}-\boldsymbol{A}\right)^{-1}\boldsymbol{E}_w$, $G(z)=\boldsymbol{C}\left(z\boldsymbol{I}-\boldsymbol{A}\right)^{-1}\boldsymbol{B}$ is the system transfer function , $\{a_i\}$ and $\{b_i\}$ are respectively the coefficients of the denominator polynomial $D(z)$ and numerator polynomial $N(z)$ of $G(z)$. Note that the equation error $\upsilon(k)$ is a sum of two colored noise processes generated by white noise processes $w(k)$ and $v(k)$.

2.1. Identification using System the Model

Let us consider the problem of identification of the system model (2) using the well-known, and widely used LS approach. Since the equation error is colored noise process $\upsilon(k)$ given by (3), the estate will not be unbiased and efficient. Let $F_w(z)$ be some whitening filter such that:

$$F_w(z)\upsilon(z) = e(z) \tag{4}$$

where $e(k)$ is a zero mean white noise process. Filtering the linear regression model (2) yields:

$$y_f(k) = -\sum_{i=1}^{n} a_i y_f(k-i) + \sum_{i=1}^{n} b_i u_f(k-i) + e(k) \tag{5}$$

where $u_f(k)$ and $y_f(k)$ are filtered input and the output $u_f(z) = F_w(z)u(z)$ and $y_f(z) = F_w(z)y(z)$. Our problem is to find this whitening filter, and is addressed in the next section on the Kalman filter.

Kalman filter: The state-space model $((\boldsymbol{A}-\boldsymbol{KC}),\boldsymbol{B},\boldsymbol{K},\boldsymbol{C})$ of the Kalman filter relating the system input $r(k)$ and system output $y(k)$ to the predicted output $\hat{y}(k)$, termed *predictor form*, is:

$$\begin{aligned} \hat{\boldsymbol{x}}(k+1) &= (\boldsymbol{A}-\boldsymbol{KC})\,\hat{\boldsymbol{x}}(k) + \boldsymbol{B}\,r(k) + \boldsymbol{K}y(k) \\ y(k) &= \boldsymbol{C}\hat{\boldsymbol{x}}(k) + e(k) \end{aligned} \tag{6}$$

The residual, or the innovation sequence, is a zero- mean white noise process given by:

$$e(k) = y(k) - \hat{y}(k) \tag{7}$$

where $F(z) = |(z\boldsymbol{I} - \boldsymbol{A} + \boldsymbol{KC})|$ is termed the *Kalman polynomial* , $\hat{\boldsymbol{x}}(k)$ and the predictor $\hat{y}(k)$ are respectively the best estimates of $\boldsymbol{x}(k)$ and $y(k)$.

Expression for the residual: The frequency-domain expression relating the inputs $r(z)$ and $y(z)$ to the residual $e(z)$ is given by (Doraiswami & Cheded, 2012)

$$e(z) = \frac{D(z)}{F(z)}y(z) - \frac{N(z)}{F(z)}r(z) \tag{8}$$

Expressing as a difference equation, yields:

$$y(k) = -\sum_{i=1}^{n} a_i y(k-i) + \sum_{i=1}^{n} b_i r(k-i) + \sum_{i=0}^{n} c_i e(k-i) \tag{9}$$

Kalman filter residual and the equation error: By comparing the expressions of the equation error term $\upsilon(z)$ for the process form and the Kalman filter residual $e(z)$, it can be deduced that:

$$\upsilon(z) = F(z)e(z) \tag{10}$$

Comparing (4) and (10) we conclude that the whitening filter $F_w(z) = \frac{1}{F(z)}$. Thus the Kalman filter is in fact the whitening filter

3. Proposed HOLS Method

Deconvolving both of the polynomials $D(z)$ and $N(z)$ by $F(z)$ yields a high order model:

$$y(k) = -\sum_{i=1}^{n_h} a_{hi} y(k-i) + \sum_{i=1}^{n_h} b_{hi} r(k-i) + e(k) \tag{11}$$

The ($2n_h$x1) feature vector $\boldsymbol{\theta}_h$ is:

$$\boldsymbol{\theta}_h = \begin{bmatrix} a_{h1} & a_{h2} & . & a_{hn_h} & b_{h1} & b_{h1} & . & b_{hn_h} \end{bmatrix}^T \tag{12}$$

Using the LS approach the estimate $\hat{\boldsymbol{\theta}}_h$ of the high- dimensional feature vector $\boldsymbol{\theta}_h$, yields:

$$\hat{\boldsymbol{\theta}}_h = \left(\boldsymbol{H}_h(k)^T \boldsymbol{H}_h(k) \right)^{-1} \boldsymbol{H}_h^T(k) \boldsymbol{y}(k) \tag{13}$$

where $\hat{\boldsymbol{\theta}}_h = \begin{bmatrix} \hat{a}_{h1} & \hat{a}_{h2} & . & \hat{a}_{hn_h} & \hat{b}_{h1} & \hat{b}_{h1} & . & \hat{b}_{hn_h} \end{bmatrix}^T$, and $\boldsymbol{H}_h(k)$ is an (Nx$2n_h$) data matrix. If the order n_h is sufficiently large, the quality of the estimate of the HOLS method will be the same as that of the LS when the equation error is a zero-mean white noise process. As the equation error is a zero-mean white noise process, the estimates will thus be unbiased and efficient.

The PEM: The linear regression model (9) is used. Let the augmented (M_ax1) vector $\boldsymbol{\theta}_a$, formed of the coefficients where $M_a = 3n$, is given by :

$$\boldsymbol{\theta}_a = \begin{bmatrix} a_1 & . & a_n & b_1 & . & b_n & c_1 & . & c_n \end{bmatrix}^T \tag{14}$$

The PEM estimate is obtained from a nonlinear optimization problem and its objective is to minimize the quadratic function of the prediction error $V_N = \frac{1}{2}\sum_{k=1}^{N} e^2(k)$.The PEM estimates are obtained recursively using the Newton-Raphson method.

SM method: The SM uses an approach similar to the high-order method in that the impulse response sequence is truncated to some finite but large number of terms.

4. Illustrative Example:

Higher order model and reduced order model derivations

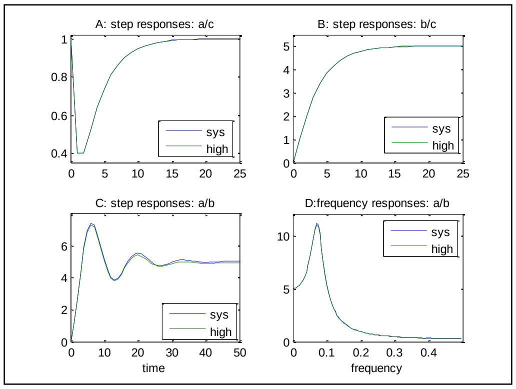

Fig. 1. The step and frequency responses of HOLS.

High order model approximation:

Let $D(z)=1-1.6z^{-1}+0.8z^{-2}$ and $N(z)=z^{-1}$ and $F(z)=1-z^{-1}+0.2z^{-2}$. where $v(k)$ and $w(k)$ are both zero mean white noise processes with unit variance, The linear regression model is given by:

$$
\begin{aligned}
&y(k)=1.6y(k-1)-0.8y(k-2)+r(k-1)\\
&+e(k)-e(k-1)+0.2e(k-2)
\end{aligned}
\tag{15}
$$

Using deconvolution and selecting the order $n_h=7$ yields:

$$\frac{D(z)}{F(z)}=\frac{1-1.6z^{-1}+0.81z^{-1}}{1-z^{-1}+0.2z^{-2}}\approx\sum_{i=1}^{7}a_{hi}z^{-i} \text{ and } \frac{N(z)}{F(z)}=\frac{z^{-1}}{1-z^{-1}+0.2z^{-2}}\approx\sum_{i=0}^{7}b_{hi}z^{-1}$$

Fig. 1 shows the performance of the high-order system. Subfigure A compares the step response of the FIR approximation model obtained from the deconvolution of $D(z)$ by $F(z)$ with the IIR model $\frac{D(z)}{F(z)}$, while subfigure B compares the step response of the FIR approximation model obtained from the deconvolution of $B(z)$ by $F(z)$ with the IIR model $\frac{B(z)}{F(z)}$. Subfigures C and D compares respectively the step and frequency responses of both the high-order model $G_h(z)$ and the original system model $G(z)$.

Reduced order model: We will pretend that the correct model order is not known and has to determined. Consider the model given by (15) . The variance of the zero-mean white noise process was unity. The number of data samples was $N=1024$. The high-order model was chosen to be a 7^{th}-order linear regression model. The reduced- order models were derived from the high-order model for the selected orders of 2,3,4 and 5 . These 4 reduced-order models were analyzed using the AIC measure. The second-order model was found to have the minimal AIC measure among the 4 reduced-order models. The identified second-order model $\hat{G}(z)$ is given by:

$$\hat{G}(z)=\frac{-0.0578+1.070z^{-1}+0.0055z^{-2}}{1-1.603z^{-1}+0.8050z^{-2}} \tag{16}$$

The identified model was very close to the true model: $\hat{G}(z)=\dfrac{1.070z^{-1}}{1-1.6z^{-1}+0.81z^{-2}}$.

The performance of the two-stage identification used in selecting a 7^{th}-order high-order model and reducing it to the selected orders of 2, 3, 4 and 5 is shown in Fig. 2. Subfigure A shows the step responses of the system, the high-order model, and all the four reduced order models, while subfigure B gives the AIC measure for each of the 4 model orders. Subfigures C, D, E and F show the pole-zero maps of the 4 reduced-order models.

Comments: The step responses of the system, the high-order model and of all the four reduced-order models are almost identical. Note that the order of the high-order model should be low enough to ensure that the data matrix is well conditioned, yet large enough to ensure that the equation error is whitened, indicating that the high-order model has captured completely the dynamic behaviour of the system. The condition number for the 7^{th} order high order model was 21.05.

5. Comparison of the PEM, the HOLS and the LS

The performances of the identification using the LS method, the HOLS, and the PEM were compared. The state-space model (1) is given by:

$$\begin{aligned} \boldsymbol{x}(k+1) &= \begin{bmatrix} 1.6 & 1 \\ -0.8 & 0 \end{bmatrix} \boldsymbol{x}(k) + \begin{bmatrix} 1 \\ 0 \end{bmatrix} r(k) + \begin{bmatrix} 1 \\ 0 \end{bmatrix} w(k) \\ y(k) &= \begin{bmatrix} 1 & 0 \end{bmatrix} \boldsymbol{x}(k) + v(k) \end{aligned} \tag{17}$$

where $v(k)$ and $w(k)$ are both zero-mean, unit-variance white noise processes, $D(z) = 1 - 1.6z^{-1} + 0.8z^{-2}$ and $N(z) = z^{-1}$. The poles are $\lambda(A) = 0.8 \pm j0.4$.

Prediction error method

The Kalman gain $\boldsymbol{K} = \begin{bmatrix} 0.9175 \\ -0.5661 \end{bmatrix}$, Kalman polynomial $F(z) = 1 + c_1 z^{-1} + c_2 z^{-2}$, $c_1 = -0.6825$, $c_2 = 0.2339$. The poles are $\lambda(\boldsymbol{A} - \boldsymbol{KC}) = 0.3412 \pm j0.3428$

$$\begin{aligned} \hat{\boldsymbol{x}}(k+1) &= \begin{bmatrix} 0.6825 & 1 \\ -0.2339 & 0 \end{bmatrix} \hat{\boldsymbol{x}}(k) + \begin{bmatrix} 1 \\ 0 \end{bmatrix} r(k) + \begin{bmatrix} 0.9175 \\ -0.5661 \end{bmatrix} y(k) \\ \hat{y}(k) &= \begin{bmatrix} 1 & 0 \end{bmatrix} \hat{\boldsymbol{x}}(k) \end{aligned} \tag{18}$$

Identification model for the LS method: The feature vector and the data vector are

$$\begin{aligned} \boldsymbol{\theta} &= \begin{bmatrix} -1.6 & 0.8 & 1 \end{bmatrix}^T \\ \boldsymbol{\psi}^T(k) &= \begin{bmatrix} -y(k-1) & -y(k-2) & r(k-1) \end{bmatrix} \end{aligned} \tag{19}$$

Identification model for the PEM:

$$e(z) = \frac{\left(1 - 1.6z^{-1} + 0.8z^{-2}\right) y(z)}{1 - 0.6825z^{-1} + 0.2339z^{-2}} - \frac{z^{-1} r(z)}{1 - 0.6825z^{-1} + 0.2339z^{-2}} \tag{20}$$

Identification model for HOLS: The high-order model derived in the previous section was used.

Table 1 shows the quality of the estimates using the LS, the HOLS and the PEM methods. The estimates $\{\hat{a}_i\}$ and $\{\hat{b}_i\}$ of the plant parameter coefficients $\{a_i\}$ and $\{b_i\}$ are given when the true values were $a_1 = -1.6$, $a_2 = 0.8$ and $b_1 = 1$.

Table 1.The LS, HOLS and PEM coefficients.

	$[1 \quad \hat{a}_1 \quad \hat{a}_2]$	$[\hat{b}_0 \quad \hat{b}_1 \quad \hat{b}_2]$
LS	$[1 \quad -1.2886 \quad 0.5029]$	$[0 \quad 1.0069 \quad 0]$
HOLS	$[1 \quad -1.5999 \quad 0.7929]$	$[-0.034 \quad 1.03 \quad -0.03]$
PE	$[1 \quad -1.6015 \quad 0.7984]$	$[0 \quad 1.0060 \quad -0.0313]$

Fig. 3 shows the performance of the LS, HOLS and PEM methods. Subfigures A, B and C show respectively the step responses of the system identified using the LS, HOLS and PEM methods. The true step response of the system is also shown for comparison.

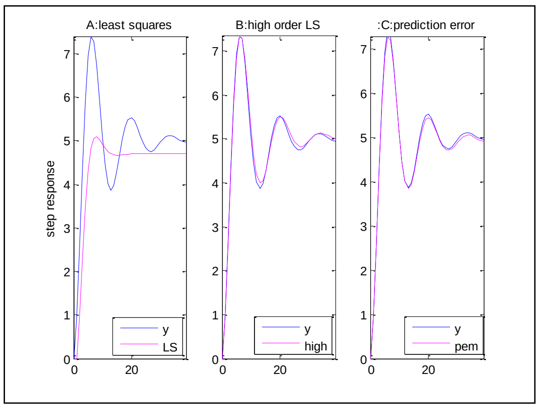

Fig. 3. Comparison of the step responses: LS, HOLS and PEM methods.

Comments: The equation (8) ties both the PEM and HOLS methods as the equation error is a white noise process. In the PEM $\|e\|^2$ is minimized but this is achieved at the expense of having the coefficients a_i, b_i and c_i enter nonlinearly the expression of $e(k)$. Moreover, with PEM, the parameter estimation problem has no closed-form solution and is solved recursively using some gradient-based methods with possible initialization and local minima problems and slow convergence rates. The HOLS method may then be used as it is offers a closed-form of the estimates of the unknown parameters, is computationally efficient despite the fact that number of parameters to be estimated is large. In practice however, with only a moderately high order, the HOLS performance approaches that of the PEM. In our experience, a high model order of about thrice the true one is sufficient.

6. Evaluation on a Physical Two Tank Process Control System

The two tank process control system is formed of two tanks connected by a pipe. The leakage is simulated in the tank by opening the drain valve. A Direct Current (DC) motor-driven pump supplies the fluid to the first tank and a Proportional Integral (PI) controller is used to control the fluid level in the second tank by maintaining the level at a specified level, as shown in Figure 4.

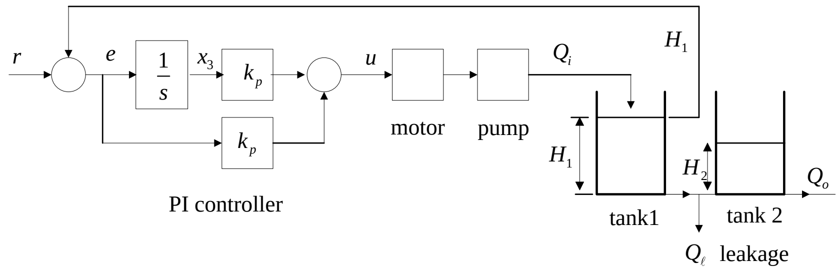

Fig. 4. Two tank process control system.

Consider the Figure 10.7, where $r(k)$, $e(k)$, $u(k)$, $H_1(k)$ and $H_2(k)$ are respectively the reference input, error input driving the controller, control input to the DC motor, height of the first tank, height of the second tank; Q_i, Q_o and Q_ℓ are respectively the (volumetric) flow rate of the inflow from the pump to the first tank , outflow from the second tank, and leakage outflow from the pipe connecting the two tanks; A_1 and A_2 are respectively the cross-sectional areas of the first and the second tanks. The data from the process control system, namely the flow rate, the height and the control input, are acquired using LABVIEW interfaced to a personal computer (PC). The objective is to identify the transfer functions of the three subsystems $G_i(z): i = 0,1,2$, which are defined by $G_0(z) = G_{eu}(z)$ relating the error $e(k)$ and the control input $u(k)$, $G_1(z) = G_{uq}(z)$, relating the control input $u(k)$ and the flow rate $q_i(k)$, and $G_2(z) = G_{qh}(z)$, relating the flow rate $q_i(k)$ and the height $h_2(k)$. The high order least squares method was used. Since the dynamics of the physical system contain uncertainty in the form of unmodeled dynamics including nonlinear effects such as saturation of the flow rate as shown in subfigure C in Figure 5, a high order model of order 15 was used. Using the model order reduction, a 3rd was model was derived (Doraiswami, 2005).

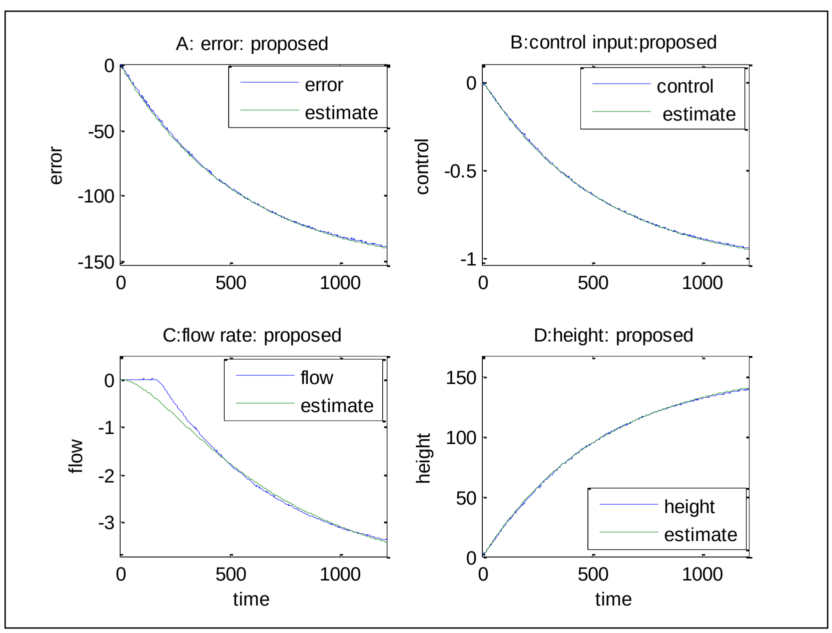

Fig. 5. The error, control input, flow rate and the height, and their estimates.

7. Conclusion

It is relatively simple to use the HOLS method to identify a model based on the Kalman filter (KF) structure. As the equation error is a zero-mean white noise process, the HOLS estimates of the parameters of the high-order model is unbiased and satisfies the Cramer Rao lower band (i.e. is efficient). The order of the high-order model should be selected such that it is low enough to ensure that the data matrix is well conditioned, yet large enough to ensure that the high-order model has captured completely the dynamic behaviour of the system. The higher the selected order, the larger is the condition number. The HOLS is computationally efficient and yields a closed form solution to the parameter estimation problem, and its performance is very close to that of the PEM.

Acknowledgement

The first author acknowledges the support of the ECE Department, University of New Brunswick, and The NSERC of Canada. The second author acknowledges the support of KFUPM, Saudi Arabia.

References

Doraiswami, R. (2005). A two-stage Identification with Application to Control, Feature Extraction , and Spectral Estimation. *IEE Proceedings: Control Theory and Applications, 152*(4), 379-386.

Doraiswami, R., & Cheded, L. (2012). Kalman Filter for Fault Detection: an Internal Model approach. *IET Control Theory and Applications*, *6*(5), 1-11.

Doraiswami, R., & Cheded, L. (2013). A Unified Approach to Detection and Isolation of Parametric Faults Using a Kalman Filter Residuals. *Journal of Franklin Institute, 350*(5), 938-965.

Doraiswami, R., & Cheded, L. (2013). Fault Diagnosis of a Sensor Network:A Distributed filtering Approach. *Transactions of the ASME, Journal of Dynamic Systems, Measurement, and Control, 135*, 1-10.

Forssell, U., & Ljung, L. (1999). Closed Loop Identification Revisited. *Automatica*, *35*(7), 1215-1241.

Kumaresan, R., & Tufts, D. W. (1982). Estimating Exponentially Damped Sinusoids and Pole-Zero Mapping. *IEEE Transactions on Acoustics, Speech and Signal Processimng, 30*(6).

Ljung, L. (1999). *System Identification: Theory for the User*. New Jersey: Prentice-Hall.

Mendel, J. (1995). *Lessons in Estimation Theory in Signal Processing, Communications and Control.* Prentice-Hall.