Application of an Improved LWR Method to Real-Time Aircraft Parameter Identification Problems

Giampiero Campa*, Marcello Napolitano**, Brad Seanor***, Mario Luca Fravolini**, Yongkyu Song*

+ Department of Electronic and Information Engineering, Perugia University, 06100 Perugia, Italy.
* Department of Mechanical and Aerospace Engineering, West Virginia University, Morgantown, WV 26506/6106.
- School of Aerospace and Mechanical Engineering Hankuk Aviation University, Goyangshi, S. Korea

Abstract

In this paper a recently introduced time domain-based parameter identification (PID) method is considered for the estimation of the aerodynamic coefficients of a specific aircraft. The basic method is the Local Weighted Regression (LWR) technique and it is improved through the addition of a data Retention-and-Deletion (RD) strategy. The resulting PID algorithm was tested using NASA F/A-18 HARV flight data for evaluating its capabilities for on-line real time PID purposes. The results were then compared with theoretical estimates from wind tunnel analysis and with the PID results from the Maximum Likelihood method, a well known PID method for off-line applications.

List of Acronyms

HARV High Alpha Research Vehicle
LWR Locally Weighted Regression
LS Least Squares
ML Maximum Likelihood
PID Parameter Identification
RD Retention & Deletion
SVD Singular Value Decomposition
WT Wind Tunnel

1. Introduction

Aircraft parameter identification (PID) from flight data has been extensively conducted as a post flight analysis for several years. Several statistical methods have been used for PID purposes with the Maximum Likelihood method being one of the most widely used approaches. In recent years drastic increases in the available on-board computational power have allowed to consider on-line real time applications of PID techniques. In particular, the on-line extension of the PID process has immediate and potentially very important applications for control of time varying aircraft systems, such as an aircraft subjected to substantial changes in dynamic and aerodynamic characteristics. A fast convergence of the parameters to be estimated is clearly a critical point for this type of application. Most of the available on-line PID algorithms are based on variations of the Least Squares (LS) algorithm, such as Recursive Least Square (RLS) 2, Real-time Batch Least Squares (BLS) 3, and Extended Kalman Filtering (EKF) 4. The real time applications of any of these methods presents a substantial challenge due to a combination of the unavoidable presence of system and measurement noise, the lack of information for PID purposes in the flight data (such as a prolonged steady state flight condition), and potential unavailability of independent control inputs – a necessary condition for accurate PID – due to the interactions with the closed-loop control laws. Analytical approaches to handle some of the above problems include the use of temporal and spatial constraints (such as forgetting factors and/or the use of short set of flight data).

In trying to overcome some or all the problems described above, this effort focuses on the interface between a recently introduced PID method, called the Locally Weighted Regression (LWR) and a Retention-Deletion technique applied to the PID data. The resulting method has been applied to the estimation of the aerodynamic coefficients of the NASA F/A-18 HARV aircraft model. The results are compared with the estimations from the Maximum Likelihood method and values from wind tunnel analysis.

2. LWR Problem

Multiple linear regression analysis is one of the most widely used approaches for the estimation of a vector of parameters from a collection of “almost-linearly-related” input-output data. This approach is based on linear algebra and leads to an elegant formulation and a straightforward analysis, allowing the use of powerful and very well known algorithms. The reliability of these methods comes from the property that a pseudo-inverse solution for a linear system with more equations than unknown coefficients is optimal in the least squares sense. The general linear regression model is given by:

\[ Y = X_\beta + \epsilon \]  \hspace{1cm} (1)

where \( Y \) is a \((n \times 1)\) vector of known responses of the system, \( X \) is a \((n \times p)\) matrix of known inputs to the system (note that the last column of this matrix is usually a column of ones
allowing for a “bias”, namely a constant input to the system, to be introduced, \( \beta \) in the \((p \times 1)\) vector of parameters to be estimated, and \( \varepsilon \) is a \((n \times 1)\) vector of independent normal random variables, with zero mean (\( E(\varepsilon) = 0 \)) and unknown diagonal variance-covariance matrix. This matrix is generally assumed to be a multiple of the \((n \times n)\) identity matrix: \( \sigma^2 I \). Therefore we have that \( E(Y) = X\beta \) and \( \sigma^2 \{Y\} = \sigma^2 I \). The problem is to find the vector \( \beta \) such that \( X\beta \) (which is the expected value of \( Y \)) is as close as possible (in the least squares sense) to \( Y \), so that \( \sigma^2 \) is minimized. The objective is to find the value of \( \beta \) that minimizes the following quadratic index:

\[
Q = \varepsilon^T \varepsilon = (Y - X\beta)^T (Y - X\beta)
\]

The solution to this problem is given by:

\[
b = (X^T X)^{-1} X^T Y
\]

It can be shown - using the Gauss-Markov theorem - that this solution is such that the error vector:

\[
e = Y - Xb
\]

has zero mean - meaning unbiased estimation - and minimum variance among all the possible linear unbiased solutions. Furthermore, it can be shown that the resulting estimation for \( \sigma^2 \) is the MSE value (Mean Square Error):

\[
MSE = \frac{\varepsilon^T \varepsilon}{n - p}
\]

The covariance of the solution is:

\[
\sigma^2 \{b\} = E\{(b - E\{b\}) (b - E\{b\})^T\} = (X^T X)^{-1} \sigma^2 \{Y\} X (X^T X)^{-1} = \sigma^2 (X^T X)^{-1}
\]

Substituting the MSE in lieu of \( \sigma^2 \) in equation (6) we obtain:

\[
\sigma^2 \{b\} = \frac{\varepsilon^T \varepsilon}{n - p} (X^T X)^{-1}
\]

This can be used as an on-line reliability measure for the estimated parameters.

The Local Weighted Regression (LWR) is a particular multiple linear regression analysis method which is characterized by the weighting of the involved linear system by a diagonal matrix. This weighting is implemented before the actual computation of the pseudo-inverse. The weighting allows selecting the “most important” data while, at the same time, avoiding to completely forgetting the past information enclosed in the rest of the available data. It is obvious that this property could be precious in on-line identification problems where the system to be estimated is time varying.

In the LWR method the equation (1) is weighted with a diagonal matrix \( W \) to express the major/minor importance of a particular row of data. Therefore:

\[
WY = WX\beta + W\varepsilon
\]

The goal is to find the value of \( \beta \) minimizing the following quadratic index:

\[
Q = \varepsilon^T W^T \varepsilon = (WY - WX\beta)^T (WY - WX\beta)
\]

The solution is provided by:

\[
b = (WX)^T WY = (X^T W^2 X)^{-1} X^T W^2 Y
\]

where \( W^2 = W^T W \). This solution is formulated so that the weighted error vector:

\[
W\varepsilon = WY - WX\beta
\]

has zero mean – implying unbiased estimates – and minimum variance among all the possible linear unbiased solutions. The covariance matrix of the solution is given by:

\[
\sigma^2 \{b\} = E\{(b - E\{b\}) (b - E\{b\})^T\} = (X^T W^2 X)^{-1} X^T W^2 \sigma^2 \{Y\} W^2 X (X^T W^2 X)^{-1}
\]

Substituting the MSE in lieu of \( \sigma^2 \) in equation:

\[
\sigma^2 \{b\} = \frac{\varepsilon^T \varepsilon}{n - p} (X^T W^2 X)^{-1} X^T W^2 X (X^T W^2 X)^{-1}
\]

Thus, if \( W = I \), we can also consider \( X^T W^2 X (X^T W^2 X)^{-1} I = I \).

### 3. The Retention and Deletion Problem

For simplicity purposes, \( D \) denotes the data matrix, with the provision that if a diagonal weighting is used (as in the LWR method) then \( D = W \), otherwise \( D = X \). For on-line real time PID problems, the size of \( D \) is limited by the available computational power. Therefore, once the number of rows in \( D \) reaches a pre-defined value, the issue of which rows in \( D \) should be replaced by the new data arises. A simple and immediate strategy would be to delete the oldest rows, or the data that are in the least squares sense “most different” from the current data. However, this strategy could lead to the deletion of rows containing precious information for the identification process, causing, therefore, ill conditioning. This, in turn, increases the variances of the estimation error associated with the parameters to be estimated. A more appropriate strategy would be to delete the rows of \( D \) that “bring less information” for the identification process. Thus, when those rows are replaced, it is likely that the new rows of data will bring more information than the ones just deleted. This, in turn, would also decrease the variances for the identified parameters. This strategy is accomplished by replacing the rows of \( D \) whose deletion cause the trace of \( (D' D)^{-1} \) to increase the least, since the lower is this trace, the lower is the variance of the estimated parameters, as shown in (13).

Assume that the current data matrix is given by:

\[
D = \begin{bmatrix} X \\ H \end{bmatrix}
\]

Assume that the row \( H \) should be deleted. Before the deletion, the trace of the following matrix is needed:

\[
P^{-1} = (D' D)^{-1} = (X^T X + H^T H)^{-1}
\]

The trace of \( P^{-1} \) is related to the trace of:

\[
(X^T X)^{-1} = R^{-1}
\]

From the definitions of \( P \) and \( R \):

\[
R = P - H^T H
\]

\[
198
\]
Inverting $R$ and applying the Matrix Inversion Lemma the following expression is generated:

$$R^{-1} = (P - H^T H)^{-1} P^{-1} = P^{-1} - P^{-1} H^T (H P^{-1} H^T - I)^{-1} H P^{-1}$$

(18)

Thus:

$$tr(R^{-1}) = tr(P^{-1}) - tr(P^{-1} H^T (H P^{-1} H^T - I)^{-1} H P^{-1})$$

(19)

Since $H P^{-1} H^T$ is scalar:

$$tr(P^{-1} H^T (H P^{-1} H^T - I)^{-1} H P^{-1}) = \frac{tr(P^{-1} H^T H P^{-1})}{H P^{-1} H^T - I}$$

(20)

$$= \frac{tr(H P^{-1} H^T)}{H P^{-1} H^T - I} = V V^T$$

where $V V^T$ is the singular value decomposition of $P$, (so $V V^T = I$ and $S$ is diagonal). Thus the problem reduces itself to find the particular row $H$ between all the rows of $D$ that maximizes (20). Since $HV$ is just the $i$-th row of the matrix $B=DV$, and since $S$ is diagonal:

$$tr(P^{-1} H^T (H P^{-1} H^T - I)^{-1} H P^{-1}) = \sum_{j=1}^{n} b_j^2 \frac{s_j}{s_j} - 1 = F(i)$$

(21)

Therefore, the row $H$ in the matrix $D$ maximizing (20) will have index $k = \max_i F(i)$.

4. Implementation Aspects of the Algorithm

The resulting PID algorithm is essentially a merger of the LWR method with the Retention & Deletion (RD) scheme outlined above. Within $D=WX$, the weight matrix $W$ varies with time according to the following algorithm. Let $h$ be the index of the “newest” row in $X$—that is the one just inserted. The distance between the current row and the others can then be computed. Next, define a column vector $d$ such that its entry $d_i$ is the “distance” between the $h$-th row and the $i$-th one:

$$d_i = \sum_{j=1}^{n} (x_{ij} - x_{hj})^2$$

(22)

The diagonal matrix $W$ can be set as:

$$W_{ii} = e^{-d_i / \Delta t^2}$$

(23)

One of the characteristics of the used algorithm is that $k$, the “gaussian window width”, is not a constant, but a discrete state variable, which starts from an initial value and evolves according to the following equations:

$$\sigma_m^2 = \max_{t \in [0, T]} \sigma_m^2$$

(24)

$$\Delta k = \max(-0.025, \min(\sigma_m^2 - \sigma_m^2, 0.025))$$

$$k(t + 1) = \max(0.5, \min(k(t) + \Delta k, 20))$$

In other words, if $m$ is the index of the estimated parameter having the maximum variance, its variance can be expressed in percentage versus a predefined “maximum variance” $\sigma^2$ for the $m$-th parameter. Therefore, if this percentage is high $k$ increases and $W$ becomes closer to the Unity Matrix. Otherwise $k$ decreases and $W$ becomes more “selective” putting more weight in the data in the rows that are “similar” to the most recent data. Note that the absolute value of the increment cannot be more than 0.025, and the Gaussian window width is set in the range $[0.5, 20]$. For the application of the RD scheme a few issues need to be addressed. In fact, although the parameters and the variances of the relative estimates are calculated using the equations (10) and (13) respectively, no weight is considered for what concerns the retention and deletion algorithm. Therefore, the data matrix is $X$ in lieu of $WX$. Furthermore, although equations (10) estimates $p$ parameters, in general only a subset of the $p$ parameters might be required. In that case, since only a subset of $p$ is needed, only the “partial trace” of $(X^T X)^d$ is considered, that is the sum of the diagonal elements of $(X^T X)^d$ corresponding to the parameters that are actually sought. The resulting RD will feature the following steps:

1. Compute the partial trace of $(X^T X)^d$.
2. If the partial trace is less than half the number of parameters actually sought, it can be assumed that the inverse matrix has been calculated without numerical problems. Thus, the algorithm simply deletes the rows that are furthest from the current one.
3. If the partial trace is more than half the number of derivatives sought, then delete the rows causing the partial trace of $(X^T X)^d$ to increase the least, using (21) with $s_j = \infty$ if the $i$-th parameter is not of interest for PID purposes.

In terms of required computational effort, it should be noticed that the same singular value decomposition (SVD) scheme used to select the rows (s) to be deleted through (21) can be used to perform the inversion of the data matrix needed for the on-line evaluation of both estimates (10) and its variances (13). Therefore, the algorithm is substantially reduced to a single SVD for each time step. Real time performance has been reached by running a non-optimized Matlab code implementation of the algorithm on a Pentium 300 Processor.

5. NASA F/A-18 HARV Dynamics

The F-18 HARV aircraft dynamics is modeled in the standard spherical coordinate form $(\alpha, \beta, \phi, \psi)$ of the body-axis equations of motion:

$$V_t = \frac{1}{m} [-D \cos \beta + Y \sin \beta + T \cos \alpha \cos \beta]$$

$$-g(\cos \alpha \cos \beta \sin \theta - \sin \beta \sin \phi \cos \theta)$$

$$-\sin \alpha \cos \beta \cos \phi \cos \theta$$

$$-\sin \alpha \sin \beta + q - \tan \beta (p \cos \alpha + r \sin \alpha)$$

(25)

$$\alpha = \frac{1}{m V_t \cos \beta} [-L - T \sin \alpha] + \frac{g}{V_t \cos \beta} (\cos \alpha \cos \phi \cos \theta)$$

$$+ \sin \alpha \sin \theta + q - \tan \beta (p \cos \alpha + r \sin \alpha)$$

(26)
\[ \dot{\beta} = \rho \sin \alpha - r \cos \alpha + \frac{1}{mV_T} (D \sin \beta + \frac{Y}{V_T} - T \cos \alpha \cos \beta) + \frac{8}{V_T} (\cos \alpha \sin \beta \sin \theta + \cos \alpha \sin \phi \cos \theta - \sin \alpha \sin \beta \cos \phi \cos \theta) \]

\[ l_1 \dot{q} + l_2 \dot{r} + (l_1 - l_2) p q - I_m q = \dot{L} \]

\[ l_1 \ddot{q} + (l_1 - l_2) p \dot{q} + I_m (p^2 - q^2) = M + (d_{1Y} - x_g) \cdot T \cdot C_{m_{\delta}} \ddot{\psi} \]

\[ l_1 \ddot{r} - l_2 p \dot{r} + (l_1 - l_2) p q + I_m q r = N + (d_{1n} - x_g) \cdot T \cdot C_{m_{\delta}} \ddot{\phi} \]

where \( V_T \) is the true airspeed, \( \alpha, \beta \) are the attack and sideslip angles, \( p, q, r \) are the angular rates around the axes \( x, y, z \), and \( \theta \) are the Euler roll and pitch angles. The command inputs \( \delta_1, \delta_2, \delta_3, \delta_4, \delta_5, \delta_6, \delta_7, \delta_8 \) are respectively the elevator, aileron, rudder, leading edge flap, trailing edge flap, symmetric aileron, pitch vane, and differential horizontal tail deflections. The aerodynamic forces and moments are expressed as

\[ N = C_{n_{\delta}} \frac{\bar{q}}{\bar{S}} + (d_{1n} - x_g) \frac{T}{q_S} C_{m_{\delta}} \]

\[ M = C_{m_{\delta}} \frac{\bar{q}}{\bar{S}} + (d_{1n} - x_g) \frac{T}{q_S} C_{m_{\delta}} \]

\[ \beta_i = \frac{x_1}{b} \cdot \frac{T}{q_S} C_{m_{\delta}} \]

where \( x_1 = [\alpha \ \delta_1 \ \delta_2 \ \delta_3 \ \delta_4 \ \delta_5 \ \delta_6 \ \delta_7 \ \delta_8] \)

\[ x_2 = [\beta \ \delta_1 \ \delta_2 \ \delta_3 \ \delta_4 \ \delta_5 \ \delta_6 \ \delta_7 \ \delta_8] \]

The aerodynamic forces acting on an aircraft can be expressed in terms of the aerodynamic coefficients, \( C_L, C_D, C_m, C_{\alpha}, C_{\theta}, C_n \). Therefore the aerodynamic coefficients derivatives with respect to the state and input variables are then used to evaluate the coefficients of the \( A \) and \( B \) matrices of the linearized model around a certain operating point \( \bar{x}_0 \). When flight conditions change, a reliable on-line estimator should identify the respective changes in the aerodynamic derivatives as accurately as possible to allow, for example, the computation of a new controller at new operating conditions.

The first order approximation of a given aerodynamic coefficients \( C \) can be expressed as:

\[ C = C_0 + \frac{\partial C}{\partial \psi} \psi \]

where \( C_0 \) is a constant and \( \psi \) is the column vector of (input and state) variables functionally related to the coefficient \( C \).

Defining:

\[ x = [\psi^T \ 1] \]

\[ \beta = \left[ \frac{\partial C}{\partial \psi} \right] C_0 \]

\[ \left[ \begin{array}{c}
C(t_1) \\
C(t_2) \\
\vdots \\
C(t_n)
\end{array} \right] = \left[ \begin{array}{c}
x(t_1) \\
x(t_2) \\
\vdots \\
x(t_n)
\end{array} \right] \]

The estimation problem can then be set up exactly as described in (1):

\[ Y = \dot{X} \beta + \epsilon \]

Thus, the derivatives vector \( \beta \) can be estimated using the approach outlined above.

For the specific F-18 HARV dynamics, there is a total of 47 parameters to be estimated. The estimation requires the mass, the moments of inertia, as well as the airspeed and the altitude at the given flight conditions.

\[ C = x \beta \]
7. Results

A detailed analysis was conducted to evaluate the performance of the proposed on-line parameters identification scheme (LWR-RD), in particular a comparison was carried out with the well-known Maximum Likelihood (ML) off-line method. Furthermore, the results have been compared with available wind tunnel (WT) values for the NASA F/A-18 HARV aircraft. The analysis was conducted for both the longitudinal and the lateral-directional dynamics, with PID maneuvers conducted at angles of attack (\(\alpha\)) ranging from 20 to 40 degrees. The results of the time histories of the estimates for selected longitudinal and lateral directional coefficients are shown in Figures 1-4 and Figures 5-7 respectively. Note the estimates for the first 2.5 seconds are not reported due to the initial large transient associated with the LWR/RD algorithm.
Also note the wind tunnel estimates for $c_{L_{a}}$ and $c_{m_{o}}$ are not available due to inconsistencies between the mathematical model in the simulation code and the standard aerodynamic model of the aircraft. In fact, in the aerodynamic model of the NASA F/A-18 HARV aircraft developed by NASA Dryden researchers, the bias terms of the aerodynamic force and moment coefficients are embedded in the $\alpha$-derivatives. Overall the performance of the LWR-RD method are satisfactory with the estimates converging within 15-20 sec and with reasonable deviations with respect to the values from off-line parameter identification methods.

8. Conclusions

The paper describes the results of an effort focused on improving a recently introduced PID algorithm, the Locally Weighted Regression (LWR) for on-line real time applications. The algorithm was enhanced with a Retention & Deletion (RD) scheme allowing the use of meaningful data for PID purposes at any given time instant. The results were compared with both the estimation obtained from wind tunnel data experiments and results provided by the Maximum Likelihood method. The so-modified (LWR + RD) technique provided fairly accurate estimates for both lateral and longitudinal aerodynamic coefficients. Furthermore, the LWR and RD techniques share the same SVD routine. This allows a substantial optimization of the required computational effort. The performance of the (LWR + RD) method, both in terms of accuracy of the estimates and rate of convergence, along with its reasonable computational effort, make the approach suitable for on-line applications.

Acknowledgement

Partial support for the 1st, 2nd, and 3rd authors has been provided through a grant from the Institute for Software Research (ISR), Fairmont, WV. Partial support for the 4th and 5th authors has been provided by grants respectively from the Italian National Research Council (CNR), and the Korea Science and Engineering Foundation (Grant Number 1999-1-305-001-5).

References