A simple method for bias reduction in time domain least squares parameter estimation

James Welsh, Graham Goodwin, Hugues Garnier

To cite this version:


HAL Id: hal-00161991
https://hal.archives-ouvertes.fr/hal-00161991
Submitted on 12 Jul 2007
A SIMPLE METHOD FOR BIAS REDUCTION IN TIME DOMAIN LEAST SQUARES PARAMETER ESTIMATION

James S. Welsh * Graham C. Goodwin * Hugues Garnier **

* EE & CS, The University of Newcastle, Australia 2308.
** Centre de Recherche en Automatique de Nancy CNRS UMR 7039, Université Henri Poincaré, Nancy, France.

Abstract:
This paper examines the issue of bias that arises, in particular when applying a traditional continuous-time model identification technique to resonant systems. We propose a method to reduce the impact of this bias by pre-processing the data. This new method requires no knowledge of the noise colouring. An example is presented that shows the superior performance of the proposed method over that of a traditional method.

Keywords: Least Squares, Time Domain, Parameter Estimation

1. INTRODUCTION

Due to its simplicity and a reasonably fast rate of convergence, the least squares method of parameter estimation is widely adopted in system identification (Hsia, 1977; Strejc, 1980; Zhao et al., 1992). However, it is well known (Goodwin and Payne, 1977; Stoica and Söderström, 1982) that the least squares parameter estimation method is generally biased and hence doesn’t provide a consistent estimate of the parameters when the input and/or output signal is contaminated with noise.

A number of methods exist which compensate for the bias in the estimates (Sagara and Wada, 1977; Stoica and Söderström, 1982; Sakrison, 1967). Early methods of bias compensation were primarily based on subtracting the bias from the least squares algorithm. (Sakrison, 1967) proposed a method based on stochastic approximation to correct a mean-square equation error criterion utilising knowledge of the input/output noise variances. A methodology similar to this was described in (Furuta and Paquet, 1970) where the noise variances themselves were unknown but their ratio assumed known. In contrast to this a recursive scheme (Sagara and Wada, 1977) was developed such that when the output was corrupted by unknown additive noise a consistent estimator is obtained by recursively estimating the noise variance and utilising it in a recursive least squares algorithm. This scheme is commonly referred to as Compensated Least Squares. Asymptotic accuracy of this algorithm was comprehensively studied in (Stoica and Söderström, 1982).

A multi-step bias compensating least squares identification method for continuous time systems has also been developed (Zhao et al., 1992). In this method the plant model is augmented with a pre-filter, with distinct characteristic roots. Parameters are then estimated for the augmented system. A linear transformation is next performed to obtain the desired model parameters. Parameter estimation accuracy in this method depends significantly on a heuristic choice of the pre-filter. To avoid this somewhat arbitrary choice of pre-filter it was proposed (Nguyen et al., 1993) to utilise the filter provided by the classical generalised least squares method. Further to this it was established (Stoica et al., 1995) that these methods actually belong to a class of weighted instrumental variable estimators.
More recently an algorithm known as direct Bias-
Eliminated Least-Squares (BELS) was proposed (Zheng,
1998). This method again is a multistep procedure
which does not involve a pre-filtering step. Instead it
estimates the cross covariance directly and uses this
estimate to obtain an unbiased parameter estimate.
Once again it has been shown that this algorithm is
equivalent to a simple instrumental variables estimator
with the instruments being delayed inputs (Söderström
et al., 1999). Further refinement of the BELS method,
when applied specifically to ARMAX model structures,
has been examined and the relationship with IV
methods explored (Zheng, 2004).

It has been recently suggested that bias may be par-
ticular bad when applied to resonant systems. From a
practical point of view these particular systems are
found in many applications. For example, many
mechatronic systems exhibit resonant behaviour (see
for example Akcay and Ninness (1999), Moheimani
(2000)). Control of such systems typically relies upon
the availability of high precision models. Now, in
principle, such models could be obtained from phe-
nomenological considerations but usually this type of
model is too complex to be accurately described using
physical parameters. Hence, one needs to obtain
reliable models using data collected from the system.

Pre-filtering of data for use in parameter estimation is
not a new concept as seen from the above discussion.
It has also been used in the past to focus the parameter
estimate over a desired bandwidth (Ljung, 1999).

This paper is aimed at reducing the bias associated
with the estimation of continuous-time systems using
least squares. In particular we develop this method for
a specific set of systems, namely resonant systems.
Bias is often analysed by looking at correlations be-
tween regressors and noise. However, a better idea
may be to simply use the asymptotic expression for
the least squares cost function. We explore this idea
here. We further assume that a deterministic input is
applied to the open loop system consisting of energy
at specific frequencies only. In contrast to the previous
mentioned methods we then develop a relatively sim-
ple method to reduce the bias based on pre-filtering
which is based on knowledge of the input frequencies.

We use Monte Carlo simulations to highlight the effect-
iveness of the proposed filters on the least squares
estimator and contrast the performance with a simple
instrumental variables estimator.

The remainder of the paper is organised in the fol-
lowing way. In section 2 we outline a simple method
of identifying the coefficients of differential equations
using time domain data. Section 3 examines the po-
tential source of bias in the estimate. In section 4 we
show how the effect of bias can be reduced by ap-
propriate pre-filtering. In section 5 we provide simulation
results which show the effectiveness of the proposed
filtering and demonstrate that the method presented
here has performance comparable to a simple instru-
mental variables method. Finally in section 6 we draw
conclusions.

2. IDENTIFICATION IN THE TIME DOMAIN

Consider a single-input single-output linear time-
invariant causal system whose input \( u(t) \) and
output \( y(t) \) are related by a constant coefficient
differential equation of order \( n \),
\[
g^{(n)}(t) + a_{n-1}g^{(n-1)}(t) + \ldots + a_0g(0) = 

b_nu^{(m)}(t) + \ldots + b_0u(0) 
\tag{1}
\]
where \( x^{(i)}(t) \) denotes the \( i \)th time-derivative of the
continuous-time signal \( x(t) \).

Equation (1) can be written as
\[
A(p)y(t) = B(p)u(t),
\tag{2}
\]
with
\[
B(p) = b_mp^n + \cdots + b_1p + b_0,
\]
\[
A(p) = p^n + a_{n-1}p^{n-1} + \cdots + a_1p + a_0, \quad n \geq m,
\]
where \( p \) is the differential operator, i.e. \( px(t) = \frac{dx(t)}{dt} \).

The polynomials \( A(p) \) and \( B(p) \) are assumed to be
relatively prime and the roots of the polynomial \( A(p) \)
are assumed to have negative real parts, hence the
system under study is asymptotically stable. It is also
assumed that the continuous-time signals \( u(t) \) and
\( y(t) \) are sampled at a regular time interval, \( T_s \).
The sampled signals are denoted as \( \{u(t_k); y(t_k)\} \).

Furthermore the system is assumed to be subject to an
arbitrary set of initial conditions
\[
u_0 = \left[ u(0) \ u^{(1)}(0) \ldots u^{(m-1)}(0) \right], \\
y_0 = \left[ y(0) \ y^{(1)}(0) \ldots y^{(n-1)}(0) \right].
\]

The identification problem is to then estimate the
coefficients of the differential equation model from
\( N \) sampled measurements of the input and output,
\[
\mathcal{Z} = \{u(t_k); y(t_k)\}_{k=1}^{N}.
\]

The literature essentially describes two time-domain
approaches to estimate a continuous-time model from
discrete-time data. The first is to estimate from the
sampled data, a discrete-time model and then convert
it into a continuous-time model. The second consists of
identifying directly a continuous-time model from
the discrete-time data. In this paper we consider the
second approach. We note that by taking this approach
the differential equation model is not a linear combi-
nation of the sampled process input and output signals,
i.e. it contains time-derivative terms which are not
available as measurement data in most practical cases.

Various types of methods have been devised to deal
with the need to reconstruct these time-derivatives.
One early approach (Young, 1964) is known as the
state variable filter (SVF) method. Due to the simplicity of this method we utilise it here for the purpose of demonstrating the benefits of the proposed method of bias reduction.

2.1 The State Variable Filter Approach

Consider the Laplace transform of the differential equation as defined in (1),

\[ A(s)Y(s) = B(s)U(s) + C(s), \]

with

\[ C(s) = c_{n-1}s^{n-1} + \cdots + c_1s + c_0 \]

(4)

where \( s \) represents the Laplace variable and \( Y(s) \) and \( U(s) \) are the Laplace transforms of \( y(t) \) and \( u(t) \) respectively. The coefficients \( c_i \) depend on the unknown parameters \( a_i \) and \( b_i \) as well as the unknown initial conditions. Assume that a filter has a Laplace transform \( L(s) = 1/E(s) \) where all the zeros of \( E(s) \) lie in the left half plane. Applying this filter to both sides of (3) yields

\[ \frac{A(s)}{E(s)}Y(s) = \frac{B(s)}{E(s)}U(s) + \frac{C(s)}{E(s)}. \]  

or

\[ \frac{1}{E(s)}Y(s) + \sum_{i=0}^{n-1} a_i \frac{s^i}{E(s)} Y(s) = \sum_{i=0}^{m} b_i \frac{s^i}{E(s)} U(s) + \sum_{i=0}^{n-1} c_i \frac{s^i}{E(s)}. \]  

(6)

The minimum-order SVF filter is typically chosen to have the following form

\[ L(s) = \frac{1}{E(s)} = \left( \frac{p_n}{s + p_n} \right)^n \]  

(7)

where \( p_n \) is the breakpoint frequency. This latter quantity can be chosen in order to emphasize the frequency band of interest and it is advised, in general, to choose it slightly larger than the bandwidth of the system to be identified (Young, 1964). Let \( L_k(s) \), for \( k = 0, 1, 2, \ldots, n \), be a set of filters defined as

\[ L_k(s) = \frac{s^k}{E(s)} = \left( \frac{p_n}{s + p_n} \right)^k \]  

(8)

and \( l_k(t) \) be their corresponding functions in the time-domain. By using the filters defined in (8), equation (6) can be rewritten as

\[ \left( L_n(s) + a_{n-1}L_{n-1}(s) + \cdots + a_0L_0(s) \right)Y(s) = \left( b_nL_n(s) + \cdots + b_0L_0(s) \right)U(s) + \left( c_{n-1}L_{n-1}(s) + \cdots + c_0L_0(s) \right). \]  

(9)

In terms of time-domain signals, (9) can be written as

\[ [L_0y](t) = b_0[L_0u](t) + \cdots + b_0[L_0y](t) + c_{0}L_0(t) \]

\[ [L_{n-1}y](t) + \sum_{i=1}^{n-1} a_i[L_{n-i-1}y](t) + \cdots + a_0[L_0y](t) = b_n[L_nu](t) + \cdots + b_0[L_{n-1}u](t) + c_{n-1}L_{n-1}(t) + \cdots + c_0L_0(t) \]  

(10)

where

\[ [L_iy](t) = l_i(t) \ast y(t) \]

\[ [L_iu](t) = l_i(t) \ast u(t) \]

and \( \ast \) denotes the convolution operator. The filter outputs \([L_iy]\) and \([L_iu]\) provide the time-derivatives of the inputs and outputs in the frequency band of interest. These may then be exploited for linear regression and other parameter estimation techniques.

At time-instant \( t = t_k \), equation (10) can be rewritten in standard linear regression form as

\[ [L_ny](t_k) = \phi^T(t_k) \theta \]  

(11)

where

\[ \phi^T(t_k) = [-[L_{n-1}y](t_k) \ldots -[L_0y](t_k)] [L_{m}u](t_k) \ldots [L_0u](t_k) [l_{n-1}(t_k) \ldots l_0(t_k)] \]

\[ \theta = [a_{n-1} \ldots a_0 \ b_m \ldots \ b_0 \ c_{n-1} \ldots c_0]^T. \]  

(12)

Now, from \( N \) samples of the input and output signals observed at discrete times \( t_1, \ldots, t_N \), the least-squares (LS)-based SVF estimates are given by

\[ \hat{\theta}_N = \left[ \sum_{k=1}^{N} \phi(t_k)\phi^T(t_k) \right]^{-1} \sum_{k=1}^{N} \phi(t_k) [L_ny](t_k). \]

It is also noted that the SVF technique can also be associated with a basic instrumental variable (IV) method when the output signal is contaminated with noise (Garnier et al., 2003b).

The SVF approach also makes it possible to estimate the initial condition terms \( c_i \) along with the model parameters. However, treating them as an additional set of unknowns does add complexity to the parameter estimation. From (10), it can be seen that although the initial condition terms do not vanish, the impulse responses of the low-pass filters decay exponentially and hence become insignificant quite quickly. Thus, if the SVF-based algorithm is used with a large observation time \( T \), the terms related to the initial conditions may be neglected after a time \( T_0 = b_0T \). The parameter estimation algorithm is then applied over \([T_0, T]\), where \( T_0 \) is chosen comparable to the settling time of the filter (7). The number of parameters to be estimated can, in this way, be reduced substantially which is advantageous with regard to computation effort and numerical properties.

Since only the sampled versions of the continuous-time signals are available, the output of the state-variable filters is computed from a discrete approximation. This problem is well known and should be treated in a proper manner since errors generated by the digital implementation can have a significant influence on the quality of the estimated model (Chou et al., 1999). Using a control canonical form, the state-space representations of the continuous-time SVF filter can be either integrated by the Runge-Kutta method or discretized by using an appropriate method provided the
From Parseval’s Theorem, the cost function in (16) treating this problem we proceed as follows. Firstly, we expand the second term in (19) as,

$$\frac{\hat{A}(j\omega)}{E(j\omega)}H_o(j\omega) = \left(1 + \hat{\Lambda}(j\omega)\right).$$  (20)

Equation (19) then becomes,

$$J(\theta) = \int_{-\pi}^{\pi} \left| \frac{\hat{A}(j\omega)}{E(j\omega)}G_o(j\omega) - \frac{\hat{B}(j\omega)}{E(j\omega)} \right|^2 \phi_{uu}(\omega) \, d\omega \quad + 2\pi d_n + d_n \int_{-\pi}^{\pi} \left| \hat{\Lambda}(j\omega) \right|^2 \, d\omega$$  (21)

or more simply,

$$J(\theta) = \int_{-\pi}^{\pi} \left( \frac{\hat{A}(j\omega)}{E(j\omega)}G_o(j\omega) - \frac{\hat{B}(j\omega)}{E(j\omega)} \right)^2 \phi_{uu}(\omega) \, d\omega \quad \quad \quad + 2\pi d_n + d_n \int_{-\pi}^{\pi} \left| \hat{\Lambda}(j\omega) \right|^2 \, d\omega.$$  (22)

It is then obvious that if we can choose

$$\frac{E(s)}{A(s)} = H_o(s),$$  (23)

then irrespective of the presence of noise, the cost function will minimise a weighted error between $G_o$ and the parameter estimate $\hat{A}$.

The presence of the $\left| \frac{1}{n} \right|^2$ in this weighting is interesting.

Our principle concern here will be with resonant systems. Therefore the estimate $\frac{\hat{B}}{\hat{A}}$ might take the form,

$$\hat{G}(s) = \sum_{i=1}^{n} \frac{2\xi_i \omega_i s}{s^2 + 2\xi_i \omega_i s + \omega_i^2}.$$  (24)

A key point to note here is that if $\hat{A} \simeq A_o$, then

$$\hat{A}(s) = \prod_{i=1}^{n} \left( s^2 + 2\xi_i \omega_i s + \omega_i^2 \right).$$  (25)

Now obviously, this is very small at the resonant peaks. Therefore if undermodelling exists, the fit will be quite insensitive to the model at these resonant peaks.
4. PRE-PROCESSING TO MITIGATE BIAS ERRORS

The second term on the right hand side of (19) indicates that if we could make \( \frac{d}{dt} \approx 0 \), then we would remove the noise term related to the colouring of the noise and hence eliminate bias.

When identifying a system in practice, a common situation is that \( \phi_n(\omega) \) will be ‘periodic’ or composed of a multi-sine test signal. In this case, there is no penalty in eliminating the noise that is not “near” an input frequency. In particular, we propose to pre-filter \( y(t) \) and \( u(t) \) with bandpass filters having their centre frequencies at the input test signal frequencies.

Say that the input frequencies are \( f_1, \ldots, f_m \) (rad/sec). Then a suitable choice for the pre-filter would be,

\[
F(s) = \sum_{i=1}^{m} \frac{2\xi_if_is}{s^2 + 2\xi_if_is + f_i^2}.
\]

We next examine different aspects of the proposed pre-filter in a heuristic sense.

The transient time of the filter \( F(s) \) is of the order of \( 1/\xi \). Hence,

(i) We want \( \xi \) ‘large’ to minimise the filter transient, but

(ii) The noise that ‘leaks’ through the filter will be of the order of \( \xi \times \text{noise spectral density at } f_i \). This suggests we should set \( \xi \) as small as possible.

(iii) Obviously the data length sets a lower limit on \( \xi \) in view of point (i).

4.1 Application of the Pre-filter

We next examine the effect of pre-filtering the data with the proposed filters on the cost of the least squares estimator.

Filtering the input/output data with the pre-filter given in (26) allows us to approximate the cost function (22) as follows,

\[
\hat{J} \approx \sum_{i=1}^{m} \left| \frac{A_i}{E_i} \right|^2 G_{oi} \left( \frac{\hat{B}_i}{\hat{A}_i} \right)^2 P_i + \left| \frac{A_i}{E_i} \right|^2 \left( \xi_i \right)^2 S_i
\]

(27)

where \( S_i \) is the noise spectral density in the vicinity of \( f_i \) and \( P_i \) is the input energy at \( f_i \).

As expected, we see that the bias is only a function of the ‘system, model and noise’ at the test signal frequencies. Indeed, we see that all we need do is to ensure that \( P_i >> \left( \xi_i \right)^2 S_i \) to reduce bias due to noise.

4.2 Rapprochement with Frequency Domain

Of course, the above idea finds a direct link to the frequency domain. Indeed, Fourier Analysis simply correlates \( \{y_k\} \) and \( \{u_k\} \) with \( \cos f_i t \) and \( \sin f_i t \). This is basically what is achieved by the filter (26). Hence, this work simply implies we can capture one of the advantages of the frequency domain when working in the time domain.

5. SIMULATION EXAMPLES

In this section we verify the results from section 4.1. Specifically we show that the proposed method, although simple, works well to reduce the effect of bias on the estimate obtained using a least squares estimator. In particular we use the standard least squares state variable filter (Issvf) method found in the The CONtinuous-Time System IDENTification (CONTSID) toolbox (Garnier et al., 2003a). We also show that the results obtained are, at least, as good as those obtained using an instrumental variables estimator.

We consider the true plant to be

\[
G_o = \frac{0.4s}{s^2 + 0.4s + 4},
\]

(28)

and excite the system using a multisine signal with each frequency component having the same amplitude. The chosen frequencies for the simulation are \( \omega = [0.5, 1, 2, 3] \) rad/sec. The sampling period was 0.005 seconds. Monte Carlo studies were conducted where each simulation was repeated 2000 times. The results were averaged and plotted. The signal to noise ratio (SNR) for the simulations was -20dB/decade. The effect of varying the damping (the quality factor) of the bandpass filters on the estimate was also examined.

We perform several sets of simulations. Firstly, we estimate the parameters using the Issvf method when no noise is present. We then carry out a Monte Carlo study using the Issvf method but in the presence of the noise as specified above. Finally, for different values of damping (\( \xi \in [0.001, 0.01, 0.1, 1] \)) in the bandpass pre-filters (given by equation (26) ) we observe the effect on the bias. Note that as we have 4 test frequencies we also have 4 bandpass filters. Figure 1 shows the bode plot for the case where the damping, \( \xi = 0.001 \).

We first observe from Table 1 that when noise is absent (Issvf (no noise)) then the parameter estimates are an accurate representation of the actual system parameters. With output noise present it is seen from Table 1 and the bode plots in Figures 2.5 that the Issvf method performs poorly. By pre-filtering the data it is observed that the estimates are significantly improved. In particular with the small value of damping (\( \xi = 0.001 \)) the estimates are as good as those obtained when no noise is present. This is shown in the both the results in Table 1 and Figures 2.5. For completeness the standard deviation for the parameter estimates are given in Table 2.
Table 1. Mean value of 2000 estimates

<table>
<thead>
<tr>
<th>Method</th>
<th>$\xi$</th>
<th>$b_1$</th>
<th>$b_o$</th>
<th>$a_1$</th>
<th>$a_o$</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>0.4</td>
<td>0</td>
<td>0.4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>lssvf (no noise)</td>
<td>0.400</td>
<td>0</td>
<td>0.4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>lssvf</td>
<td>0.246</td>
<td>-0.002</td>
<td>0.007</td>
<td>3.923</td>
<td></td>
</tr>
<tr>
<td>lssvf (pre-filtered)</td>
<td>0.001</td>
<td>0.398</td>
<td>-0.001</td>
<td>0.393</td>
<td>3.983</td>
</tr>
<tr>
<td>lssvf (pre-filtered)</td>
<td>0.01</td>
<td>0.375</td>
<td>-0.013</td>
<td>0.339</td>
<td>3.853</td>
</tr>
<tr>
<td>lssvf (pre-filtered)</td>
<td>0.1</td>
<td>0.299</td>
<td>-0.051</td>
<td>0.155</td>
<td>3.603</td>
</tr>
<tr>
<td>lssvf (pre-filtered)</td>
<td>1</td>
<td>0.254</td>
<td>-0.107</td>
<td>0.076</td>
<td>3.314</td>
</tr>
</tbody>
</table>

Table 2. Standard Deviation of 2000 estimates

<table>
<thead>
<tr>
<th>Method</th>
<th>$\xi$</th>
<th>$b_1$</th>
<th>$b_o$</th>
<th>$a_1$</th>
<th>$a_o$</th>
</tr>
</thead>
<tbody>
<tr>
<td>lssvf</td>
<td>0.001</td>
<td>0.044</td>
<td>0.079</td>
<td>0.009</td>
<td>0.121</td>
</tr>
<tr>
<td>lssvf filtered</td>
<td>0.001</td>
<td>0.052</td>
<td>0.075</td>
<td>0.054</td>
<td>0.086</td>
</tr>
<tr>
<td>lssvf filtered</td>
<td>0.01</td>
<td>0.058</td>
<td>0.083</td>
<td>0.055</td>
<td>0.097</td>
</tr>
<tr>
<td>lssvf filtered</td>
<td>0.1</td>
<td>0.0486</td>
<td>0.074</td>
<td>0.026</td>
<td>0.099</td>
</tr>
<tr>
<td>lssvf filtered</td>
<td>1</td>
<td>0.038</td>
<td>0.064</td>
<td>0.013</td>
<td>0.082</td>
</tr>
</tbody>
</table>

6. CONCLUSION

Using the asymptotic expressions for the least squares cost function we can see that the source of the bias in a least squares estimator is due to the colouring of the noise. In this paper we have shown that, when the input consists of multisines, it is possible to significantly reduce the bias by simply pre-filtering the data. Simulations studies have shown the method to work well and produce results that are comparable to those obtained using instrumental variables.


