The Performance of Generalized Minimum Variance System Identification

Capt. Odell R. Reynolds and Prof. M. Pachter
Air Force Institute of Technology/ENG
Department of Electrical and Computer Engineering
Wright Patterson AFB, OH 45433-7655
mpachter@afit.af.mil

Abstract

In this paper, a comparison is made between the novel Generalized Minimum Variance and established Generalized Least Squares estimation algorithms. The emphasis of the Generalized Minimum Variance algorithm is on the proper treatment of measurement noise for dynamical system identification. The algorithms are compared in carefully performed, reproducible experiments which include *measurement* noise. Differences are apparent under small measurement samples, but the two appear to produce statistically similar results under sufficient excitation.

1 Introduction

In this paper, a comparison is made between the novel Generalized Minimum Variance (GMV) and established Generalized Least Squares (GLS) identification algorithms. The experimental data consists of a finite number of input and output measurements of a discrete-time plant, which is representative of the pitch dynamics of a transport aircraft. Also, sensor noise has corrupted the measurements of the output. The GLS method is described in [1], and is basically the same as the method proposed in [2].

The paper is organized as follows. The GMV and GLS identification schemes are briefly described in Section 2. The setup of the experiment is discussed in Section 3. The performance of the proposed system identification algorithm, as well as a comparison to the GLS algorithm in [1], are then discussed in Section 4. Concluding remarks are in Section 5.

2 Time Domain Estimation

Consider the discrete time transfer function given by

$$T(z) = \frac{y(z)}{u(z)} = \frac{\sum_{i=0}^{n} b_i z^{n-i}}{\sum_{i=0}^{n} a_i z^{n-i}} \tag{1}$$

where $a_0 = 1$, and the $2n+1$ parameters $a_1 \ldots a_n$ and $b_0 \ldots b_n$ are unknown. This transfer function can be rewritten in a recursive form as

$$A(z^{-1})y_k = B(z^{-1})u_k \tag{2}$$

or

$$\begin{aligned} y_k = a_1 y_{k-1} + a_2 y_{k-2} + \cdots + a_n y_{k-n} + \\ b_0 u_k + b_1 u_{k-1} + b_2 u_{k-2} + \cdots + b_n u_{k-n} \end{aligned} \tag{3}$$

To identify the parameters of the discrete time system, one need only solve the following system of $2n+1$ equations in the a and b parameters

$$\begin{bmatrix} y_k \\ \vdots \\ y_{k+N} \end{bmatrix} = \begin{bmatrix} y_{k-1} & \cdots & y_{k-1+N} \\ \vdots & \cdots & \vdots \\ y_{k-n} & \cdots & y_{k-n+N} \\ \vdots & \cdots & \vdots \\ u_k & \cdots & u_{k+N} \\ \vdots & \cdots & \vdots \\ u_{k-n} & \cdots & u_{k-n+N} \end{bmatrix}^T \begin{bmatrix} a_1 \\ \vdots \\ a_n \\ b_0 \\ \vdots \\ b_n \end{bmatrix} \tag{4}$$

2.1 GMV ID Problem Formulation

Unfortunately, the values for y_k, and sometimes also u_k, cannot be directly measured. Rather, we can measure the noise corrupted Y_k, where the noise is assumed Gaussian:

$$Y_k = y_k + v_k, \qquad v_k = \mathcal{N}(0, \sigma_v^2) \tag{5}$$

which changes Eq. (3) to

$$\begin{aligned} Y_k - v_k = a_1(Y_{k-1} - v_{k-1}) + \cdots + a_n(Y_{k-n} - v_{k-n}) \\ +b_0 u_k + b_1 u_{k-1} + \cdots + b_n u_{k-n} \end{aligned} \tag{6}$$

or

$$Y_k = a_1 Y_{k-1} + \cdots + a_n Y_{k-n} + b_0 U_k + \cdots + b_n U_{k-n} + \tilde{v}_k \tag{7}$$

where

$$\tilde{v}_k = v_k - a_1 v_{k-1} - \cdots - a_n v_{k-n} \tag{8}$$

This expression is now in a form that can be set up in a statistical Linear Regression equation given by

$$\mathbf{z} = \mathbf{H}\boldsymbol{\theta} + \mathbf{v}, \qquad \mathbf{v} = \mathcal{N}(\mathbf{0}, \mathbf{R}), \qquad \mathbf{R} \geq 0 \tag{9}$$

where

$$\mathbf{z} = \begin{bmatrix} Y_k \\ Y_{k+1} \\ \vdots \\ Y_{k+N} \end{bmatrix}$$

$$\mathbf{H} = \begin{bmatrix} Y_{k-1} & Y_k & \cdots & Y_{k-1+N} \\ \vdots & \vdots & & \vdots \\ Y_{k-n} & Y_{k-n+1} & \cdots & Y_{k-n+N} \\ u_k & u_{k+1} & \cdots & u_{k+N} \\ \vdots & \vdots & & \vdots \\ u_{k-n} & u_{k-n+1} & \cdots & u_{k-n+N} \end{bmatrix}^T$$

$$\boldsymbol{\theta} = \begin{bmatrix} a_1 & \cdots & a_n & b_0 & b_1 & \cdots & b_n \end{bmatrix}^T$$

$$\mathbf{v} = \begin{bmatrix} \tilde{v}_k \\ \tilde{v}_{k+1} \\ \vdots \\ \tilde{v}_{k+N} \end{bmatrix}$$

2.2 LS System Identification

The Least Squares (LS) method assumes that $\mathbf{R} = \sigma^2\mathbf{I}$, and the estimate of the system above is then given by

$$\widehat{\boldsymbol{\theta}}_{LS} = \left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{z} \qquad (10)$$

and the parameter estimation error covariance is

$$\mathbf{P}_{LS} = \sigma^2\left(\mathbf{H}^T\mathbf{H}\right)^{-1} \qquad (11)$$

This estimate is incorrect for a linear regression corrupted by measurement noise because of the correlation introduced by that noise. One method that has been used to overcome this correlation is a simplified version of the Generalized Least Squares (GLS) algorithm.

2.3 GLS System Identification

In the GLS method, a whitening filter is used to reduce the correlation in the linear regression so the LS estimate becomes more accurate. For the case of output measurement noise only, the following model is correct; see, e.g., Eqs. (7) and (8):

$$A(z^{-1})y_k = B(z^{-1})u_k + A(z^{-1})e_k \qquad (12)$$

where e_k is a white Gaussian sequence. Since the LS estimate is only valid for uncorrelated white noise, an attempt to match an LS model is made by "dividing" the operator equation (12) through by $A(z^{-1})$, leading to

$$A(z^{-1})\frac{y_k}{A(z^{-1})} = B(z^{-1})\frac{u_k}{A(z^{-1})} + e_k$$

However, since the actual $A(z^{-1})$ is not known, what is actually done is to filter the input and output with the estimate $1/\widehat{A}(z^{-1})$. In other words,

$$\begin{aligned} \tilde{y}_k &= \widehat{a}_1\tilde{y}_{k-1} + \widehat{a}_2\tilde{y}_{k-2} + \cdots + \widehat{a}_n\tilde{y}_{k-n} + Y_k \\ \tilde{u}_k &= \widehat{a}_1\tilde{u}_{k-1} + \widehat{a}_2\tilde{u}_{k-2} + \cdots + \widehat{a}_n\tilde{u}_{k-n} + u_k \end{aligned}$$

The filtered output and input values, $\tilde{y}$ and $\tilde{u}$, are then substituted for Y_k and u_k in the $\mathbf{H}$ and $\mathbf{z}$ matrices, and the LS estimate is calculated. This process is repeated until the parameter estimate converges.

2.4 GMV System Identification

In contrast to the GLS method, the GMV method incorporates knowledge of the measurement noise structure in Eq. (3) to form a minimum variance type estimate. The minimum variance estimate of a parameter vector $\boldsymbol{\theta}$, as described in Eq. (9), is given by [3]

$$\widehat{\boldsymbol{\theta}}_{MV} = \left(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{z} \qquad (13)$$

and the parameter estimation error covariance is

$$\mathbf{P}_{MV} = \left(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right)^{-1} \qquad (14)$$

The $\mathbf{R}$ matrix is the expected value of $\mathbf{v}\mathbf{v}^T$, and if the noise v_k is assumed white, then the $\mathbf{R}$ matrix will be an $N \times N$ Toeplitz matrix with a non zero diagonal and n non zero off diagonal terms above and below the diagonal. Unfortunately, $\mathbf{R}$ is not known *a priori*, because in addition to the dependence on the given sensor's measurement error σ_v, it is a function of the (as yet unknown) coefficients of the system's transfer function, *i.e.*, $\mathbf{R} = \mathbf{R}(\boldsymbol{\theta})$. Therefore, we use Eq. (13) to search for a possible estimate. The Generalized Minimum Variance (GMV) estimate is given by the point $\widehat{\boldsymbol{\theta}}_{GMV}$ such that

$$\widehat{\boldsymbol{\theta}}_{GMV} = \left(\mathbf{H}^T\mathbf{R}^{-1}(\widehat{\boldsymbol{\theta}}_{GMV})\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{R}^{-1}(\widehat{\boldsymbol{\theta}}_{GMV})\mathbf{z} \qquad (15)$$

There are many different ways of searching for fixed points like $\widehat{\boldsymbol{\theta}}_{GMV}$, but the following simple iterative algorithm has been effective in finding the correct fixed point for signal-to-noise ratios approaching 0 dB. This algorithm is motivated by the iteration for fixed points of contraction mappings, for which the existence of a fixed point is guaranteed [4].

Step 1 - Set $i = 0$ and calculate an initial parameter estimate using LS.

$$\widehat{\boldsymbol{\theta}}_i = \widehat{\boldsymbol{\theta}}_0 = \left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{z}$$

Step 2 - Calculate $\mathbf{R}(\widehat{\boldsymbol{\theta}}_i) = E\left\{\mathbf{v}\mathbf{v}^T\right\}$.

Step 3 - Calculate $\widehat{\boldsymbol{\theta}}_{i+1}$ via Eq. (13).

$$\widehat{\boldsymbol{\theta}}_{i+1} = \left(\mathbf{H}^T\mathbf{R}^{-1}(\widehat{\boldsymbol{\theta}}_i)\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{R}^{-1}(\widehat{\boldsymbol{\theta}}_i)\mathbf{z}$$

Step 4 - If $\| \widehat{\boldsymbol{\theta}}_{i+1} - \widehat{\boldsymbol{\theta}}_i \|$ is less than some acceptable value, proceed to step 5. Otherwise, increment i and return to step 2.

Step 5 - Set $\widehat{\boldsymbol{\theta}}_{GMV} = \widehat{\boldsymbol{\theta}}_{i+1}$.

Step 6 - The error covariance of the estimate $\widehat{\boldsymbol{\theta}}_{GMV}$ is then given by

$$\left(\mathbf{H}^T\mathbf{R}^{-1}(\widehat{\boldsymbol{\theta}}_{GMV})\mathbf{H}\right)^{-1} \tag{16}$$

A fixed point for Eq. (15) does exist, at least for the problems of linear dynamical systems. Also, the algorithm has converged within the numerical limits of Matlab for all problems examined thus far. The number of iterations required is quite small for small noise levels, but does increase as the noise level increases.

Problems do arise as the noise levels increase, as is common in system identification problems at low signal-to-noise ratios. In this work, it is common for one other fixed point to appear at higher noise levels, and possible methods of dealing with this are discussed in [5]. However, the GLS method of identification tends to suffer at these noise levels as well.

3 Second Order System

For investigative purposes, we now restrict our attention to the general second-order dynamical system given by

$$T(s) = \frac{4.8s + 1.44}{s^2 + 0.84s + 1.44} \tag{17}$$

which is representative of an transport aircraft's elevator to pitch rate transfer function [6]. For this paper, it is assumed that the control is being passed through a zero-order hold with a 10 Hz sampling rate, making the discrete time transfer function for this plant

$$T(z) = \frac{0.4663z - 0.4525}{z^2 - 1.9056z + 0.9194} \tag{18}$$

The input to the system is the sum of two sinusoids; the minimum number of sinusoids to be persistently exciting for a second-order system with a zero. Thus, the elevator deflection:

$$u(t) = sin(3t) + sin(0.5t)$$

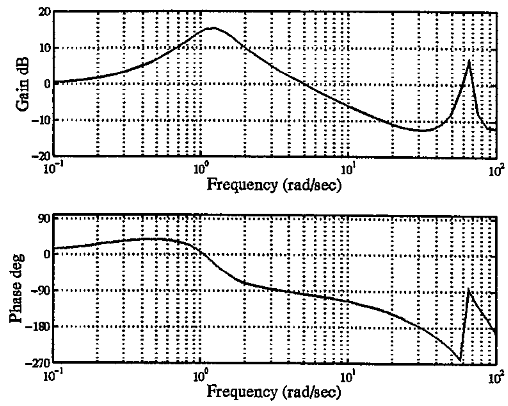

Figure 1: Discrete Bode Plot of Given Plant

In this scenario, and according to the GMV paradigm, the equation error noise vector in Eq. (8) is given by

$$\tilde{v}_k = v_k - a_1 v_{k-1} - a_2 v_{k-2}$$

To obtain the equation error covariance matrix $\mathbf{R}$, one calculates

$$E\{\tilde{v}_k \tilde{v}_{k-\tau}\} = \begin{cases} \sigma_v^2\left(1 + a_1^2 + a_2^2\right), & \tau = 0, \\ \sigma_v^2\left(a_1 a_2 - a_1\right), & \tau = 1, \\ \sigma_v^2\left(-a_2\right), & \tau = 2, \\ 0, & \tau > 2. \end{cases} \tag{19}$$

which means that the equation error covariance is the pentadiagonal matrix $\mathbf{R} = \sigma_v^2\tilde{\mathbf{R}}$, where

$$\tilde{\mathbf{R}} = \begin{bmatrix} 1 + a_1^2 + a_2^2 & -a_1 + a_1 a_2 & \cdots & 0 \\ -a_1 + a_1 a_2 & 1 + a_1^2 + a_2^2 & \cdots & 0 \\ -a_2 & -a_1 + a_1 a_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 + a_1^2 + a_2^2 \end{bmatrix}$$

4 Algorithm Comparison

In this section, the GLS and GMV estimates are compared. To do this, Matlab's [7] `randn()` function is initialized with a seed of zero and used to generate noise with a covariance $\sigma_v^2 = (0.01)^2$. This noise is then added to the true output obtained from the system shown in Fig. 1. The noise corrupted output is shown in Fig. 2.

Figure 3 displays the denominator estimates for a 100 run Monte-Carlo (MC) analysis, which upon evoking the weak law of large numbers [8], renders a gauge of the identification algorithm's estimation bias. The estimation results are first normalized by

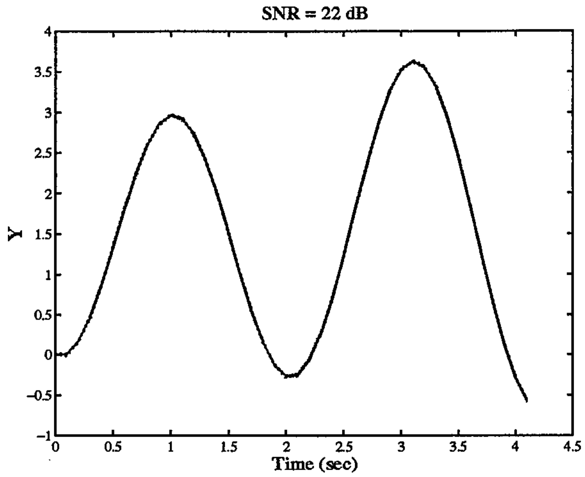

Figure 2: Plant Output with Representative Noise

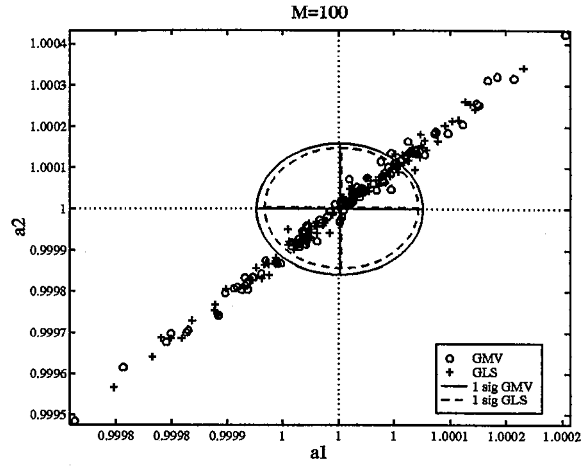

Figure 4: Denominator Estimates for 100 Measurement Linear Regression

dividing each estimate by the true estimate, and then plotted. Ellipses are plotted representing each estimation method's actual one sigma variation. The ellipses' axes are centered at the average estimate for each method. The algorithm predicted estimation error covariances are not plotted, because they are close to the actual one in each case.

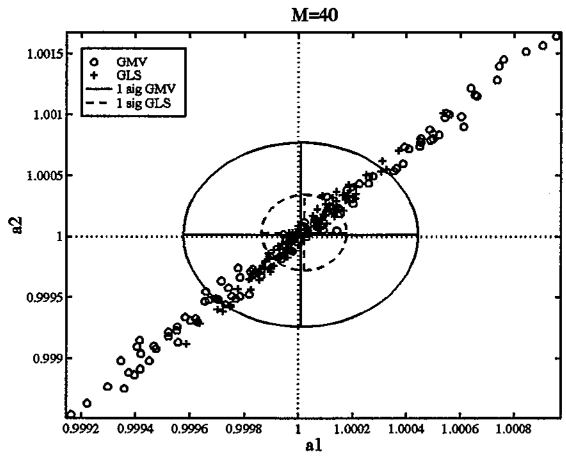

Figure 3: Denominator Estimates for 40 Measurement Linear Regression

As can be seen, the GLS estimates are quite a bit better than the GMV estimates for the case where 40 measurements are taken ($\approx$4 sec worth of data). However, this does not tell the whole story. When 100 measurements ($\approx$ 10 sec of data) are used, both estimates are better, and the statistical results are virtually the same, as seen in Fig. 4.

Another point that is not evident in Figs. 3 and 4 is the fact that the GLS method requires, in general, more measurements to obtain an accurate estimate. Figures 5 and 6 contain a representative estimation error and covariance for the GMV and GLS algorithms respectively. Each plot shows the pertinent values as measurements are added one at a time. As can be seen, the GMV algorithm produces an estimate for $M < 10$, but the GLS algorithm does not produce an estimate for all 100 noise realizations until $M = 13$.

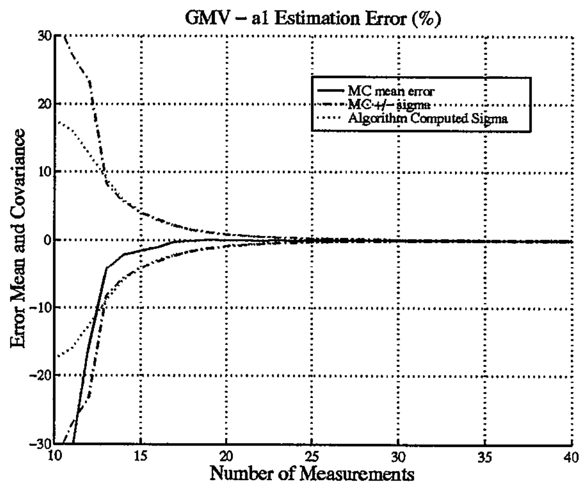

Figure 5: Estimation Error and Covariance - GMV

A scenario that causes additional trouble for the GLS method is the elimination of the initial transient data. When one allows the system to run for 10 seconds before taking data, the results in Fig. 7 are obtained. As can be seen, the two methods produce statistically similar results even for $M = 40$ here.

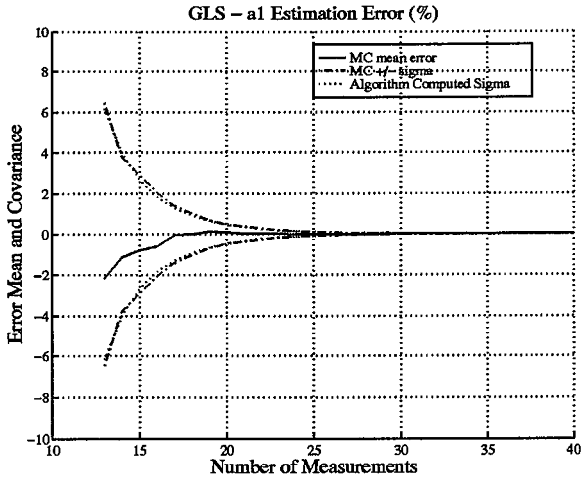

Figure 6: Estimation Error and Covariance - GLS

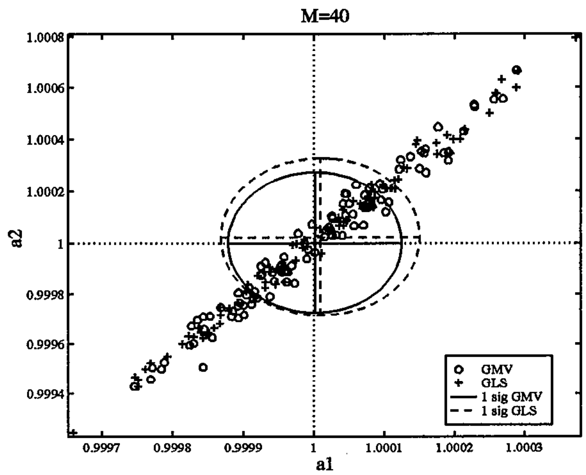

Figure 7: Denominator Estimates for 40 Measurement Linear Regression - Initial Time Removed

For this scenario, the initial differences are even more pronounced. As seen in Figs. 8 and 9, the GMV is producing accurate estimates for $M < 10$, while the GLS method is not even producing an estimate until $M = 17$. The dependence on the initial transient somewhat detracts from the GLS method's appeal, in light of the latter's operator theoretical/asymptotic argument-based derivation.

An additional advantage of the GMV method is its adaptability to other linear regression type frameworks. If there had been noise on the input as well, then the GLS derivation would have been completely incorrect. The GMV algorithm can, however, be eas-

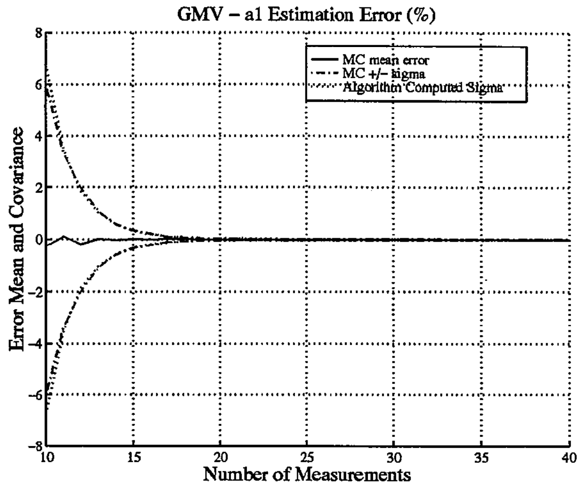

Figure 8: Without Initial Time - GMV

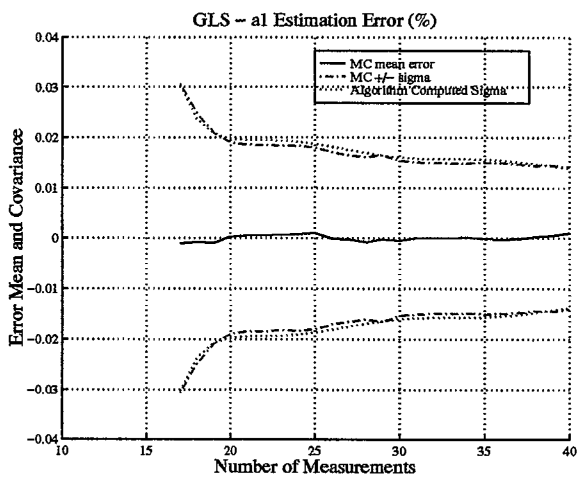

Figure 9: Without Initial Time - GLS

ily adapted to handle input noise simply by modifying $\tilde{v}_k$ to include the new noise.

5 Conclusions

In this work, the performance of the Generalized Minimum Variance and Generalized Least Squares estimation algorithms is compared. Gaussian measurement noise is assumed, as is customary in statistical filtering and system identification work. The novel minimum variance estimate equations are derived and applied to a nonlinear estimation problem. The results are then compared to the established Generalized Least Squares estimate for a second-order system that is representative of an aircraft's pitch dynamics, which is used for inner-loop flight control system design.

While the GLS method requires more measurements initially to obtain an estimate, it appears to produce a better estimate at this point than the GMV. Additionally, the GLS algorithm's performance is found to be sensitive to the initial transient in the data. However, the two methods appear to be statistically equivalent as more measurements are added. One distinct advantage of the GMV method is its ability to produce valid estimates at small measurement samples, and in general, under conditions of poorer excitation.

The GMV and GLS estimates outperformed the LS estimate in *all* cases. The LS estimate did, however, provide a useful, albeit sometimes dangerous, starting point for iterating the parameter estimate. At small noise levels, it does not matter where the algorithms are initialized, but as noise levels increase, there is an additional point of convergence that can trap the estimate. In high noise cases (low signal-to-noise ratio), it becomes necessary to examine the final result to see if it is a valid estimate. Obviously, at very low SNR, there must be at least a small amount of prior information about the plant for a valid estimate to be obtained.

In conclusion, the least squares estimate is not an effective one, even in cases of small measurement noise. It does, however, serve as a useful starting point for initializing the GLS and GMV algorithms at higher SNR. The GLS and GMV provide much more accurate estimates, and the GMV seems to provide estimates with much fewer measurements. Additionally, the GMV is more readily adaptable to differing identification requirements.

References

[1] T. J. Hsia, *System Identification.* Massachusetts: Lexington Books, 1977.

[2] K. Steiglitz and L. E. McBride, "A technique for identification of linear systems," *IEEE Transactions on Automatic Control*, vol. Ac-10, pp. 461–464, Oct. 1965.

[3] D. A. Belsley *et al.*, *Regression Diagnostics: Identifying Influential Data and Sources of Colinearity.* Wiley, 1980.

[4] T. M. Apostol, *Mathematical Analysis.* Addison-Wesley Publishing Co., 2nd ed., 1974.

[5] O. R. Reynolds, *Countering the Effects of Measurement Noise During the Identification of Dynamical Systems.* Ph.D. dissertation, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, 1996.

[6] J. H. Blakelock, *Automatic Control of Aircraft and Missiles.* John Wiley & Sons, Inc, 2nd ed., 1991.

[7] "Matlab." The MathWorks, Inc., 24 Prime Park Way, Natick, MA 01760, Dec. 1992. Version 4.2b, Copyright 1984-94.

[8] B. V. Gnedenko, *The Theory of Probability.* Chelsea Pub. Co., 1962.