Total Least Squares Estimation of Dynamical Systems

Manoranjan Majji* and John L. Junkins †

Texas A & M University, College Station, TX, 77843-3141.

The Total least squares error criterion is considered for estimation problems. Exact necessary conditions for the total error criterion have been derived. Several solution methodologies are presented to solve the modified normal equations obtained from the necessary conditions. A monte-carlo simulation depicting the effect of parameter variations on the coefficients of the resulting characteristic polynomial is performed and the approximate probability distribution of the smallest eigenvalue of interest is identified. The results are applied on the parameter identification of a novel morphing wing developed at Texas A & M University.

I. Introduction

The least squares error criterion was invented by Carl Friedrich Gauss which till to this day remains the most widely used in many diverse areas owing to its computational and statistical properties. During the turn of this century, this criterion was generalized to include uncertainty in the basis functions by Adcock.[1] Consequently, a new theory was developed called the "errors in variables" theory.[2] Golub and Van loan[3] applied this to numerical mathematics problems. Sabine Van Huffel and Van de Valle[4] applied the theory developed by Gleser to parameter estimation problems and brought the theory to systems science.

I.A. The Total Least Squares Error Criterion

The total least squares error criterion is based on generalization of the least squares error criterion. It aims at changing the basis by the slightest possible (as small as possible) to capture as much of the measurement vector as possible.

I.A.1. Paper Outline

The paper is presented as follows. First section introduces the total least squares error criterion and derives the necessary conditions in a direct manner, obtaining a modest set of nonlinear equations that are a modification of the least squares solution. Geometrical insights in to the problem are presented and a comparison is made with existing literature. Methods to solve the equations thus obtained are discussed in the next section. Two novel approaches are presented along with the celebrated SVD solution by Golub et. al,[3] Van Huffel,[4] and the well known eigenvalue problem proposed by Villegas.[5] However, at this point we consider only the developments associated with static models. Subjecting the error criterion to dynamical systems will be considered in later versions. We have to note the strong correlation of the concepts developed herein to the Minimum Model error estimation proposed by Mook and Junkins.[6]

II. Total Least Squares

As pointed out in the introduction, the least squares error criterion for minimization of residual error does not apply when linear equations are involved with an uncertainty in the basis function. Therefore, if we only have access to measurements $\tilde{A}$,of basis functions and $\tilde{y}$ of the range vector, in the linear error model

$$\tilde{y} \approx \tilde{A}x + v \tag{1}$$

*Graduate Student, Department of Aerospace Engineering, 3141 TAMU, and Student Member, AIAA.

†Distinguished Professor, Holder of Royce E. Wisenbaker Chair in Engineering, Regents Professor, Department of Aerospace Engineering, Fellow, AIAA.

where $\tilde{A} = A + V_A$ and V_A is a matrix of random variables $\sim N(0,\sigma)$ and v is the vector of random variables $\sim N(0,\sigma)$. We do not require the errors to have same statistics, however for simplicity of the derivations, we would like to impose these conditions. Straight-forward Generalizations to arbitrary statistics can be performed with arbitrary weighting as shown by Van Huffel[4] and Glesler.[2] Though the simplicity of our approach in obtaining these results makes this generalization obvious and natural, we choose not to present the general method in favor of clarity. However we do summarize the results obtained for general weight matrices in a separate section. So the Total Least Squares error criterion minimizes the cost associated with estimation errors in basis defined by $\Delta := \tilde{A} - \hat{A}$, $r = \tilde{y} - \hat{y}$

$$J = \frac{1}{2} tr(\Delta^T \Delta) + \frac{1}{2} r^T r \tag{2}$$

the necessary conditions for a minimum of these equations are given by

$$\frac{\partial J}{\partial \hat{A}} = 0, \frac{\partial J}{\partial \hat{x}} = 0 \tag{3}$$

leading to the equations (using matrix derivative identities[7]),

$$\Delta(I + \hat{x}\hat{x}^T) = (\tilde{A}\hat{x} - \tilde{y})\hat{x}^T \tag{4}$$

which, (using the Morrison Sherman Woodbury Matrix Inversion Lemma,[7]) is equivalent to

$$\hat{A} = \tilde{A} - \frac{(\tilde{A} - \tilde{y})x^T}{(1 + \hat{x}^T \hat{x})} = \tilde{A} + e\hat{x}^T \tag{5}$$

where $e := \frac{(\tilde{y} - \tilde{A}\hat{x})}{1 + \hat{x}^T \hat{x}}$, leading to the fact that the optimal correction in the data matrix, is the rank one correction $e\hat{x}^T$. The second necessity that $\frac{\partial J}{\partial \hat{x}} = 0$ yields

$$(\hat{A}^T \hat{A})\hat{x} = \hat{A}^T \tilde{y} \tag{6}$$

which, by using the expression $\hat{A} = \tilde{A} + e\hat{x}^T$ yields,

$$\left[\tilde{A}^T \tilde{A} - \frac{(\tilde{y} - \tilde{A}\hat{x})^T (\tilde{y} - \tilde{A}\hat{x}) I}{(1 + \hat{x}^T \hat{x})} \right] \hat{x} = \tilde{A}^T \tilde{y} \tag{7}$$

which we call the modified normal equations. We found that these equations are same as the ones derived by Van Huffel.[4] Substituting the necessary conditions in to the cost function, we get the optimal cost to be

$$J_1 = \frac{(\tilde{y} - \tilde{A}\hat{x})^T * (\tilde{y} - \tilde{A}\hat{x})}{(1 + \hat{x}^T \hat{x})} \tag{8}$$

The solution to the necessary conditions derived here is a nonlinear problem but the solution that minimizes J_1 is the eigenvector corresponding to the smallest eigenvalue of the symmetric positive definite form $T^T T = \begin{bmatrix} \tilde{A}^T \tilde{A} & \tilde{A}^T \tilde{y} \\ \tilde{y}^T \tilde{A} & \tilde{y}^T \tilde{y} \end{bmatrix}$, T being defined as $T = [\tilde{A} \ldots \tilde{y}]$. Defining an associated vector $z = [\hat{x}^T - 1]^T$ the minimum value of J_1 hence becomes (after making necessary substitutions of the necessary conditions),

$$J_1 = \frac{(\tilde{y} - \tilde{A}\hat{x})^T (\tilde{y} - \tilde{A}\hat{x})}{(1 + \hat{x}^T \hat{x})} = \frac{z^T T^T T z}{z^T z} \tag{9}$$

whose extremals are eigenvalues associated with the quadratic form $T^T T$, called the Rayleigh quotient.[8] The associated eigenvectors are the solutions of the problem and in this problem, the smallest eigenvalue and the corresponding eigenvector are the solutions.

II.A. Geometry of the problem

By a careful analysis of the necessary conditions for a solution to the total least squares problem, we can make some observation on the geometry of the problem. Consider the space of rectangular matrices together with the inner product definition $< A, B >: tr(A^T B)$. The norm derived from this inner product definition satisfies the polarization identity, as shown below

$$\|P+Q\|^2 + \|P-Q\|^2 = tr((P+Q)^T(P+Q)) + tr((P-Q)^T(P-Q)) \tag{10}$$

$$= 2tr(P^T P) + 2tr(Q^T Q) = 2(\|A\|^2 + \|B\|^2) \tag{11}$$

A result in analysis[9] states that a norm satisfies the polarization identity is if and only if it is derived from an inner product and is unique. This allows us to define orthogonality in matrix spaces. Now consider the inner products $\hat{y}^T(\tilde{y} - \hat{A}\hat{x})$ and $tr(\hat{A}^T\Delta)$. These are the inner products of the best estimates with the residual object in the corresponding space. Then,

$$\hat{y}^T(\tilde{y} - \hat{A}\hat{x}) = \hat{x}^T\hat{A}^T(\tilde{y} - \hat{A}\hat{x}) = 0 \tag{12}$$

$$tr(\hat{A}^T\Delta) = -tr(\hat{A}^T e\hat{x}^T) = -\hat{A}^T e = 0 \tag{13}$$

In the above expressions, the identity $\hat{A}^T e = 0$ has been used. This identity directly follows from the necessary conditions $\hat{A}^T(\hat{A}\hat{x} - \tilde{y}) = \hat{A}^T(\tilde{A}\hat{x} - \tilde{y} - e\hat{x}^T\hat{x}) = \hat{A}^T e = 0$. Therefore, the total least squares problem enforces a geometry and performs an orthogonal regression in both range space and the space of the basis functions.

III. Weighted Total Least Squares

Researchers have claimed equivalence of the weighted total least squares problems to an error criterion given by Golub and Van loan,[8] and Van Huffel.[4] However there is not sufficient freedom nor insightful conclusions from the resulting advantages unless there is a symbolic expression for the estimate. In this section we show that our technique yields a more general necessary condition that gives a lot of design freedom to alter the weights. So the cost function with appropriate weights is given by

$$J_w = tr(\Delta^T P\Delta) + r^T Q r \tag{14}$$

where P, Q are arbitrary positive definite weight matrices. They allow the designer to control the magnitudes of correction of the range and basis function tolerance levels. The first order necessary conditions $\frac{\partial J_w}{\partial \hat{A}} = 0$, $\frac{\partial J_w}{\partial \hat{x}} = 0$ yield

$$P\Delta + Q\Delta\hat{x}\hat{x}^T = Q(\tilde{A}\hat{x} - \tilde{y})\hat{x}^T \tag{15}$$

$$\hat{x} = (\hat{A}^T Q\hat{A})^{-1}\hat{A}^T Q\tilde{y} \tag{16}$$

Clearly, there is no obvious way of determining the "best" correction for giving an expression for $\hat{A}$. But indeed there is. We choose

$$\hat{A} = \tilde{A} + P^{-1}(Q^{-1} + P^{-1}(\hat{x}^T\hat{x}))^{-1}(\tilde{y} - \tilde{A}\hat{x})\hat{x}^T \tag{17}$$

We can easily verify that this expression for $\hat{A}$ satisfies the first necessary condition. Upon substitution in to the second condition, we get nothing similar to the nice eigenvalue problem. We observe that if the range measurements are weighted more $Q >> P$, then the best correction to $\hat{A}$ is applied and we get $\hat{A}_{Q\to\inf} = \tilde{A} + \frac{(\tilde{y}-\tilde{A}\hat{x})\hat{x}^T}{\hat{x}^T\hat{x}}$ and when $P >> Q$, we receive no correction in $\hat{A}_{P\to\inf} = \tilde{A}$, which is the least squares solution. In between, when there is equal uncertainty in both, we have $P \sim Q$, $\hat{A} = \tilde{A} + \alpha\frac{(\tilde{y}-\tilde{A}\hat{x})\hat{x}^T}{1+\hat{x}^T\hat{x}}$, which is the eigenvalue problem presented above. This design freedom via weighted total least squares solution is not present in the literature to the best knowledge of the authors. However the pay off is the lack of algorithms to solve this problem. Therefore we propose some new methods to solve the problem besides the eigenvalue iterations that may be extensible to solve the more general problem.

IV. Solution Methodologies

The necessary conditions being the eigenspace computations of a symmetric quadratic form can be computed stably using the Singular Value Decomposition.[8] Therefore we have the first algorithm.

IV.A. SVD Method

Since the matrix is symmetric, the smallest left singular vector is the solution to the problem (We recall that the eigenvectors and singular vectors span the same spaces in the case of a symmetric matrix).

SVD Approach.
Step 1 Compute SVD and save V, in $T^T T = U \Sigma V^T$
Step 2 $x_{SVD} = -V(1:n, n+1)/(V(n+1, n+1))$
Step 3 x_{SVD} is the required solution.

This algorithm is fairly robust but computationally expensive.

IV.B. Eigenvalue Problem

The solution, which also is the eigenvector corresponding to the smallest eigenvalue can be computed by inverse iterations.[8]

Rayleigh quotient iteration.
Step 1 Start with $z = [\hat{x}^T - 1]^T$. $\hat{x} = (\tilde{A}^T \tilde{A})^{-1} \tilde{A}^T \tilde{y}$.
Step 2 Solve $(T^T T - \lambda I) v^{(k)} = v^{(k-1)}$ for $k = 1, 2, \ldots$.
Step 3 $\hat{x}_{TLS} = -v^{(n)}(1:n, 1)/v^{(n)}(n+1, 1)$ is the converged solution.

The algorithm has cubic convergence and we can get 10 digits of accuracy in 3 iterations.

IV.C. Davidenko's Homotopy Method

Structure of the necessary conditions clearly indicates that their solution is "close" to the solution of the normal equations. This motivates us to explore the perturbation methods[10] to solve this problem. This method sees importance in the light of the expressions for the necessary conditions for the weighted total least squares formulation developed by the authors, where the eigenvalue problem is not "obvious" from the nonlinear equations (obtained in section III).

Davidenko's Method.
Step 1 Start with least squares solution $G(\hat{x}) = (\tilde{A}^T \tilde{A})\hat{x} - \tilde{A}^T \tilde{y} = 0$.
Step 2 Let $F(\hat{x}) = (\tilde{A}^T \tilde{A} - \lambda(\hat{x}) I)\hat{x} - \tilde{A}^T \tilde{y} = 0$ (Or the weighted version of it).
Step 3 Consider the Homotopy $H(z) = tF(z) + (1-t)G(z)$, $\forall t \in [0, 1]$.
Step 4 Integrate $\frac{dz}{dt} = -[\frac{\partial H}{\partial z}]^{-1}[\frac{\partial H}{\partial t}]$
Step 5 $\hat{x} = z(1)$ is the required estimate.

The accuracy of the solution depends on the numerical integrator and also is reasonably slow to compute.

IV.D. A QUEST type algorithm

QUEST is an attitude estimation algorithm, proposed by Shuster[11] which determines the "Best" attitude matrix for vector measurements. This algorithm receives attention owing to the possibility of its recursive implementation.[12] Our recursive (rather accumulative) formulations of the TLS problem has strong correlation with the REQUEST methodology. We will also derive some additional benefit from this algorithm and it is presented next. It is amazing to find that the exact developments carry forward to a general dimension from the three dimensional case involving QUEST computations. The result is summarized and the algorithm is presented next. In the eigenvalue problem,

$$\begin{bmatrix} \tilde{A}^T \tilde{A} & \tilde{A}^T \tilde{y} \\ \tilde{y}^T \tilde{A} & \tilde{y}^T \tilde{y} \end{bmatrix} = \begin{bmatrix} \hat{x} \\ -1 \end{bmatrix} = \lambda \begin{bmatrix} \hat{x} \\ -1 \end{bmatrix} \tag{18}$$

$$(S - \lambda I)\hat{x} = z \tag{19}$$

$$\lambda = \alpha + z^T (\lambda I - S)^{-1} z \tag{20}$$

where, $S := \tilde{A}^T \tilde{A}$ and $z := \tilde{A}^T \tilde{y}$. The fact that $(S - \lambda I)^{-1}$ can be expanded in lower powers (due to Cayley Hamilton theorem[13]) enables us to compute the $\hat{x}$ algebraically.

Generalized QUEST type algorithm.
Step 1 *Compute the characteristic polynomial associated with the matrix* $T^T T$*. Step 2* *With* $\lambda_0 = 0$*, as the initial guess, compute the smallest eigenvalue (Newtons root solver).*
Step 3 *Calculate* $\hat{x} = (S - \lambda I)^{-1} z$

The most useful pay off of this method is that it essentially quantifies the possible uncertainty in the solution and retains it to the characteristic equation of the symmetric quadratic form, rendering the smallest eigenvalue real and hence the uncertainty maps nonlinearly in to this one parameter and a montecarlo simulation would reveal the propagated uncertainty at this output. A test case was studied (treating $\tilde{A}$ and $\tilde{y}$, Gaussian) the following distribution of minimum eigenvalues was obtained. Notice that all the methods yield similar distributions, showing equivalence of the methods (fig. 1). Also, the mean thus obtained is typically as close as possible, we could get to the unknown "true", parameter values.

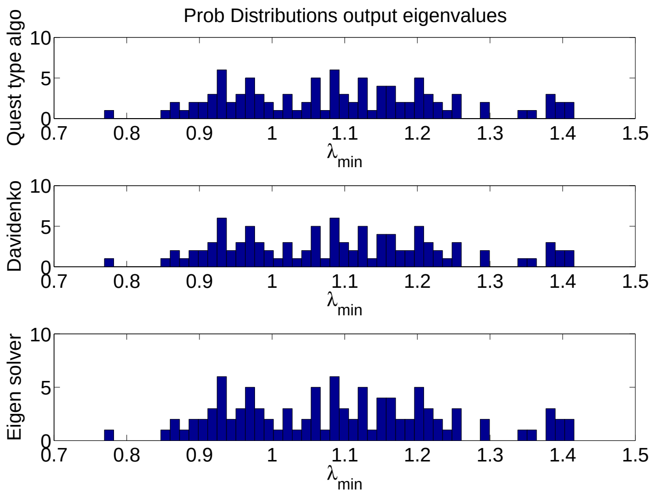

Figure 1. Probability Density Function of the minimum eigenvalue.

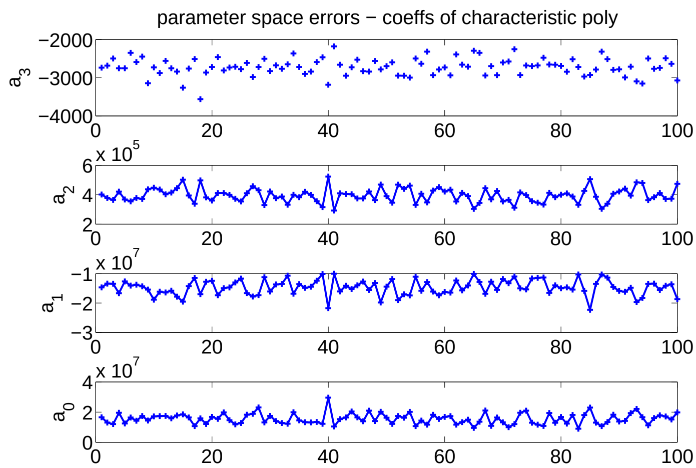

Figure 2. Uncertainty in coefficients as revealed by a Monte Carlo Simulation

But the spread is quite small especially considering the uncertainty in the magnitudes of the coefficients of the characteristic polynomial (Order of 10^2 in some, refer fig. 2). This reveals the averaging aspect of the TLS framework.

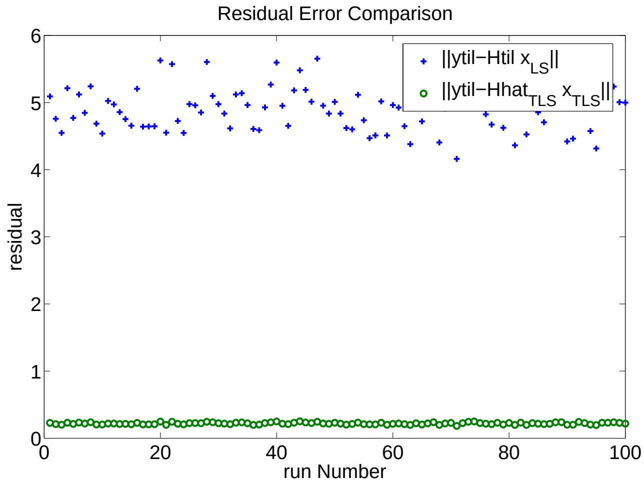

Figure 3. Residual Errors in Approximation

V. Application : Morphing Wing

The algorithms presented above were used in the identification of sensitivity coefficients of a novel morphing wing developed at Texas A& M University (fig.4). The twisting wing actuator being developed, was amenable to quasi-static aerodynamic models. As an alternative approach, we wanted to develop alternative models directly from the input output data of the wind tunnel tests. The experimental setup and aerodynamic models, along with the specifications of the tests performed are discussed in an accompanying paper.[14]

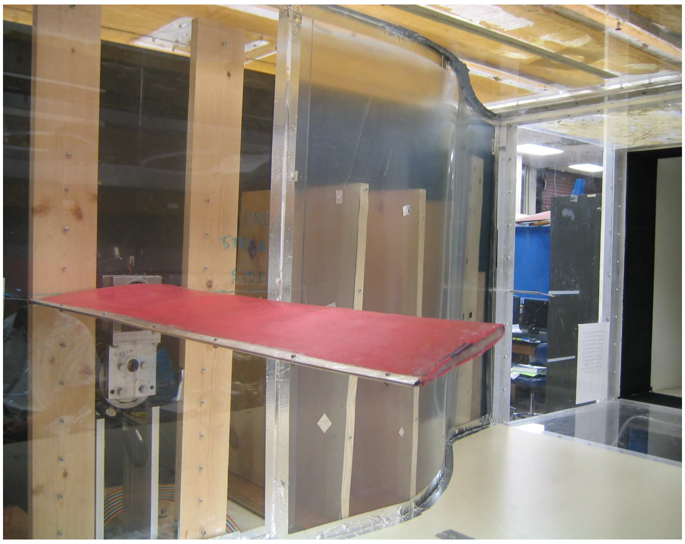

Figure 4. Morphing Wing : Experimental Setup

V.A. Discussion

The idea of using the Total least squares method as opposed to least squares method in fitting the data obtained, was to have a better approximation of the data in regions where physics based models (any strictly linear models) fail. Least squares approximation is known to "filter" the data in such regions (especially where the wing stalls) and therefore yields poor models of the physics. On the other hand, the Total least squares algorithms, possessing more knobs to turn would indeed model the physics to arbitrary extent (fig. 3) (user could control this by playing with the weights). This objective could only partially be realized because the current TLS approximation deals with "equal" magnitudes of uncertainty and thereby staying close to the least squares estimates. Upon careful observation, the better approximation of TLS is revealed by the plots 5, 6. This would be potentially increased by incorporation of weights.

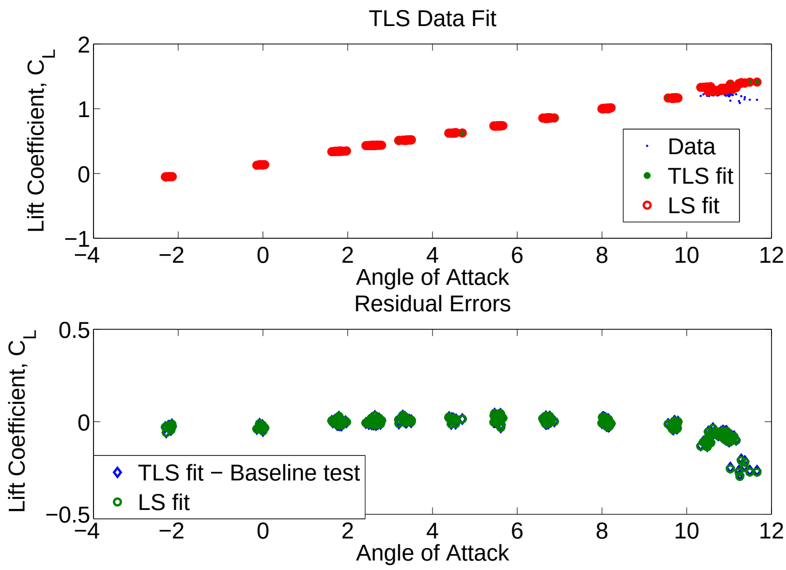

Figure 5. Baseline (no twist) test : Linear Model and Residuals

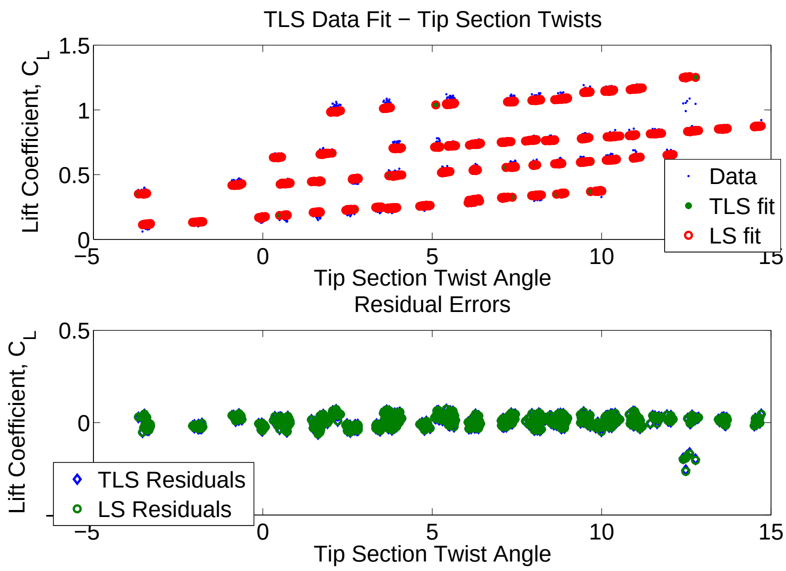

Figure 6. Tip Section Twist test: Fit of Model and Residuals

V.B. Model Validation

Once the fit was completed, a time varying test result was obtained and the prediction from least squares and the TLS method are compared in the following plots. The time varying test was performed with a small stalling and large stalling time periods and compared with the outputs of the models. It is observed that the approximation of the model fails in region where near stall conditions prevail (figs. 8, 7).

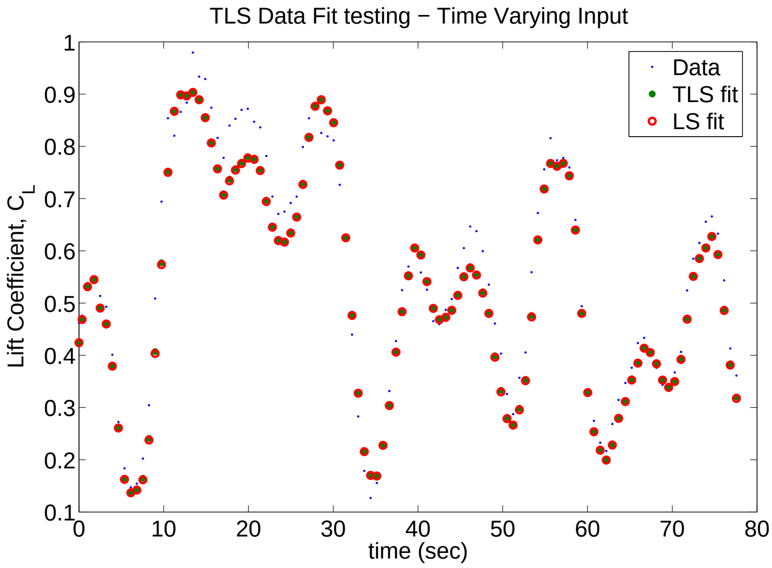

Figure 7. Model Validation : time varying input with large stalling periods

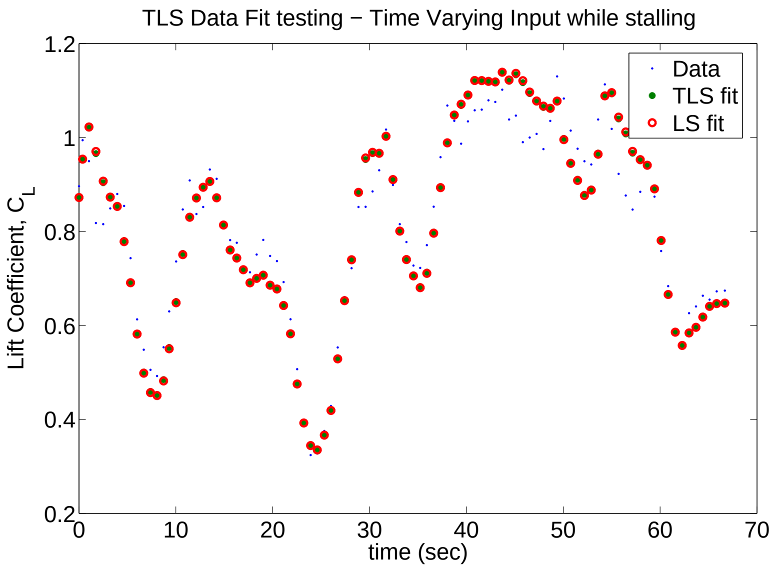

Figure 8. Model Validation : time varying inputs with small stalling periods

VI. Conclusion and Future Directions

VI.A. Conclusions

The least squares error criterion is generalized to account for errors in both range space and the basis functions in a measurement model. This was shown to lead to nonlinear necessary conditions. The necessary conditions were then realized as solutions to eigenvalue problem associated with the measurement matrix and the vector. A novel weighted total least squares criterion was presented and associated necessary conditions were derived. Several methods to solve this problem including two novel methods were presented. This was applied on a parameter identification problem of a morphing wing model developed at Texas A& M University.

VI.B. Future Directions

Most importantly because of the large degree of design choice, the method tunes itself to fit the measurements as close as possible. While in some problems, this models the physics not essentially modeled by linear least squares, in filtering problems this is not always desirable as some kind of signal reconstruction is anticipated. Therefore, work is in progress in the direction of modifying this error criterion so as to reduce the huge over parameterization. Smoother formulations incorporated for dynamical state estimation are being considered for incorporation. As mentioned in the paper, work is also in progress to develop algorithms for weighted total least squares criterion whose necessary conditions do not form the eigenvalue problem. Static problems were dealt with in the above discussion. Extension to dynamical system state estimation is expected to improved filters where there is uncertainty in the models of measurement and plant dynamics. This is being investigated and researched currently. The developments so far make optimistic gestures towards this goal.

Acknowledgments

The authors wish to acknowledge the support of Texas Institute of Intelligent Bio Nano Materials and Structures for Aerospace Vehicles funded by NASA Cooperative Agreement No. NCC-1-02038.

References

[1] Adcock, R. J., "A Problem in Least Squares," *The Analyst*, Vol. 5, Jan. – Feb. 1878, pp. 53–54.

[2] Glesler, L. J., "Estimation in a multivariable "errors in variables" regression model: large sample results," *Annals of Statistics*, Vol. 9, 1981, pp. 24–44.

[3] Golub, G. H. and Loan, C. F. V., "An Analysis of the Total Least Squares Problem," *SIAM Journal of Numerical Analysis*, Vol. 17, 1980, pp. 883–893.

[4] Huffel, S. V. and Vandewalle, J., *The Total Least Squares Problem: Computational Aspects and Analysis*, SIAM Publications, Philadelphia, 1991.

[5] Villegas, C., "Maximum Likelihood Estimation of a Linear Functional Relationship," *The Annals of Mathematical Statistics*, Vol. 32, No. 4, December, 1961, pp. 1048–1062.

[6] Mook, D. J. and Junkins, J. L., "Minimum Model Error Estimation for Poorly Modeled Dynamic Systems," *Journal of Guidance Control and Dynamics*, Vol. 3, No. 4, 1988, pp. 367–375.

[7] Crassidis, J. L. and Junkins, J. L., *Optimal Estimation of Dynamic Systems*, Chapman and Hall/CRC Press, Boca Raton, FL, 2004.

[8] Golub, G. H. and Loan, C. F. V., *Matrix Computations*, The Johns Hopkins University Press, Baltimore, MD, 3rd ed., 1996.

[9] Riesz, F. and Nagy, B. S., *Functional Analysis*, Reprinted by Dover Publications, Mineola, NY, 1990.

[10] Davidenko, D. F., "A new method of solution of a system of nonlinear equations," *Dokl. Akad. Nauk SSSR*, Vol. 88, 1953, pp. 601.

[11] Schuster, M. D. and Oh, S. D., "Three axis attitude determination from Vector Observations," *Journal of Guidance and Control*, Vol. 4, Jan. – Feb. 1981, pp. 70 – 77.

[12] Bar-Itzhack, T. Y., "REQUEST : A recursive QUEST Algorithm for Squential Attitude Determination," *Journal of Guidance Control and Dynamics*, Vol. 19, No. 5, Sep. – Oct. 1996, pp. 1034–1038.

[13] Junkins, J. L. and Kim, Y., *Introduction to Dynamics and Control of Flexible Structures*, AIAA Education Series, Washington, DC, 1991.

[14] Majji, M., Rediniotis, O. K., and Junkins, J. L., "Design of a Morphing Wing : Modeling and Experiments," Submittal, AIAA Guidance Navigation and Control Conference, 2007.