A Sampling based method for linear parameter estimation from correlated noisy measurements

Ugur Guner*. Jay H. Lee**
Matthew J. Realff***

**Georgia Institute of Technology, Atlanta, GA 30332*
USA (Tel: 404-388-2149; e-mail: Ugur.Guner@chbe.gatech.edu).
***Georgia Institute of Technology, Atlanta, GA 30332 USA (e-mail: Jay.Lee@chbe.gatech.edu)*
**** Georgia Institute of Technology, Atlanta, GA USA (e-mail: matthew.realff@chbe.gatech.edu)*

Abstract: We address the problem of linear parameter estimation in discrete time state space models in the presence of serially correlated error in variables. The common way to solve parameter estimation problem is least squares (LS) methods. LS method is not considered to be effective when both dependent and independent variables are contaminated by noise. Total Least Squares (TLS) has been introduced as the method for parameter estimation in the case of noisy response and predictor variables. However, TLS solution is not optimal when number of data is limited and noise is correlated. Constrained TLS is a variant of TLS that considers correlation of noise in the data as additional constraints. We introduced a novel method based on a stochastic sampling method to solve estimation problem from correlated noisy measurements, and we compared it with the existing methods through in silico examples. Our method demonstrates significant improvement over other common estimation algorithms, LS, TLS and Constrained TLS under the different amount of correlated noise and data points. It has the potential to be the valuable tool for the difficult real life problems, such as, biological systems where data is limited and noisy.

Keywords: Linear parameter estimation; Least Squares; Total Least Squares; Constrained Total Least Squares; Multiplicative noise; additive noise; correlation; state space models; stochastic.

1. INTRODUCTION

The problem of linear parameter estimation arises in a broad class of scientific disciplines such as signal processing, automatic control, system theory, general engineering, statistics, physics, economics, biology, medicine, etc (Huffel ,1991). Linear estimation problem becomes challenging in the presence of correlated noise. Errors are unavoidable and can be related to many sources, such as modelling, human or instruments. They may appear in different forms depending on the source of error and nature of the system. Noise can be proportional to the signal itself (multiplicative),simply additive or it can include both components. In this paper, we will particularly focus on parameter estimation in linear discrete time state space models in the presence of measurements that include both multiplicative and additive error terms. One can write the linear discrete time system as follows;

$$\hat{x}_i^{k+1} = a_{i1}\hat{x}_i^k + a_{i2}\hat{x}_2^k +,...,+ a_{iN}\hat{x}_N^k \qquad i = 1,...,N \tag{1}$$

In this equation this equation, value of observed state at time $k+1$ is linear function of all N observed states at time point, k . This model can be extended for all states and time points in a compact form as follows:

$$\hat{X}' = A^{(N\times N)}\hat{X} \tag{2}$$

$$\hat{X}'^{(N\times(M-1))} = \left\{\vec{x}^{\,2},..., \vec{x}^{\,M}\right\} \qquad \hat{X}^{(N\times(M-1))} = \left\{\vec{x}^{\,1},..., \vec{x}^{\,(M-1)}\right\}$$

where M is the number of time points and N is the number of states. Each column of X and X' is represented with the vector,

$$\vec{x}^{\,j} = \left[\hat{x}_1^{\,j},..., \hat{x}_N^{\,j}\right]^T . \tag{3}$$

The parameters are collected in matrix,

$$A^{(N\times N)} = \left\{ a_{ij} \right\}. \tag{4}$$

This paper is organized as follows. In next section we will briefly summarize common methods to solve linear estimation problem. Furthermore, we will introduce our novel approach based on a sampling algorithm. Section 3 will summarize assessment of the performance of our method compared to some common existing methods. Finally, section 4 will present the concluding remarks.

2. METHODS

2.1 Common methods

Many methods have been introduced to solve the linear estimation problem (Ljung ,1987 and Huffel ,1991). The classic way to solve the linear estimation problem is least squares. In the classical least squares regression theory, the errors are assumed to be confined only to $\hat{X}'$ (response variables) , and $\hat{X}$ (predictor variables) are assumed to be error free. One can write least squares estimate for parameters as follows,

$$A^T = \left(\hat{X}\hat{X}^T\right)^{-1}\hat{X}\hat{X}'^T \quad (5)$$

However, in our problem, it is not realistic to assume $\hat{X}$ to be error free as it shares the same columns with $\hat{X}'$ except for the first column. (See (2)). This results in serial correlation between $\hat{X}$, and $\hat{X}'$.

Total least squares (TLS) is another method of linear parameter estimation when there are errors in both sides of the equation ($\hat{X}$ and $\hat{X}'$) (Huffel ,1991).

$$\hat{X}' = X' - \Delta X' \qquad \hat{X} = X - \Delta X$$

$$\Delta X' = \left[\Delta x^2,...,\Delta x^M\right] \qquad \Delta X = \left[\Delta x^1,...,\Delta x^{M-1}\right] \quad (6)$$

where, $\Delta X'$ and ΔX are the noise terms. Since X and X' are not known, for each state ($i = 1,...,N$), equation (2) can be written in the following format;

$$x_i + \Delta x_i = a^{(1\times N)}\left(\hat{X} + \Delta\hat{X}\right)$$

(7)

where $x_i = \left[x_i^2,...,x_i^M\right]$ and $\Delta x_i = \left[\Delta x_i^2,...,\Delta x_i^M\right]$ are the i^{th} rows of $\hat{X}'$ and $\Delta X'$ respectively (See (2, 6)). $a = \left[a_{i1},..,a_{iN}\right]$ is the i^{th} row of A .

Let,

$$C = \left[\hat{X}^T \; x_i\right] \quad \text{,and} \quad \Delta C = \left[\Delta\hat{X}^T \; \Delta x_i^T\right] \quad (8)$$

Then, equation (6) can be written as;

$$(C + \Delta C)\cdot\begin{bmatrix} a^T \\ -1 \end{bmatrix} = 0$$

where a is the i^{th} row of A . The TLS problem then can be posed as follows;

$$\min\|\Delta C\|_2^F \quad \text{subject to} \quad (C + \Delta C)\cdot\begin{bmatrix} a^T \\ -1 \end{bmatrix} = 0 \quad (9)$$

The solution to this problem given as;

$$a^T = \left(\hat{X}\cdot\hat{X}^T - \lambda^2 I\right)^{-1}\hat{X}\cdot\hat{X}'^T \quad (10)$$

where λ is the smallest singular value of C . Compared to least squares solution (5), the TLS solution has a correction term, λ at the inverse of the matrix. This reduces the bias in the solution which is caused by noise in X (Kim et al, 2007). However, TLS solution inherently assumes that the noise terms, $\Delta X'$ and Δx_i are independent , which is not the case here.

The correlation between two noise term requires the total least squares solution to have additional constraints instead merely satisfying the existence of a solution (Cadzow and Wilkes, 1985). Recently, Kim et al. (2007) applied constrained least squares algorithm in the context of gene network identification problem on a linear discrete time model. In their model, they rewrite error term, ΔC in an open form and as follows;

$$\Delta C^{(M-1)\times(N+1)} = \left[\Delta\hat{X}^T \; \Delta x_i^T\right] = \begin{bmatrix} \Delta x_1^1 & ... & \Delta x_i^1 & .... & \Delta x_N^1 & \Delta x_i^2 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \Delta x_1^{M-1} & ... & \Delta x_i^{M-1} & ... & \Delta x_N^{M-1} & \Delta x_i^M \end{bmatrix} \quad (11)$$

They introduced the vector, $e_i^{(N\times 1)} = [0,..., 1,..,0]^T$ whose elements are zero except for the i^{th} element, which is equal to 1. All error terms are rewritten in a vector form as follows;

$$\Delta Y^{1\times(N(M))} = \left[\left(\Delta x^1\right)^T,...,\left(\Delta x^M\right)^T\right] \quad (12)$$

The first N columns of ΔC can be written as follows;

$$\Delta C^i = G_i\cdot\Delta Y^T \quad (13)$$

Where $G_i = \left[\left(I_{M-1}\otimes e_i\right)^T \; 0_{(M-1)\times N}\right]$ and I_{M-1} denotes the identity matrix of size $(M-1)\times(M-1)$ and the symbol, $\otimes$, denote Kronecker product of two matrices. $0_{(M-1)\times N}$ represents the matrix of zeros with size, $(M-1)\times(N)$

After several steps and simplifications, they posed this as optimization problem as follows;

$$\min\|\Delta C\|^2 = \min_a [a \;\; -1]\, C^T\left(H_a^{-1}\right)^T\left(H_a^{-1}\right) C\begin{bmatrix} a^T \\ -1 \end{bmatrix} \quad (14)$$

where $H_a = \sum_i a_i\cdot G_i$. This is a nonlinear, non-convex optimization problem without constraints. They initialized the optimization problem with least squares solution.

2.2 Our method

In this paper, we introduced a new method to solve linear parameter estimation problem when both sides of the equation are contaminated by error. This method based on a stochastic sampling approach. In this method, observations $(\hat{X}, \hat{X}')$ are perturbed by adding a negative noise term in each sample.

$$^{(j)}y_i^k = \hat{x}_i^k - {}^{(j)}\varepsilon_i^k \qquad {}^{(j)}\varepsilon_i^k \propto N\left(0, \sigma_i^k\right) \tag{15}$$

$$i = 1,\ldots,N \qquad j = 1,\ldots,S \qquad k = 1,\ldots,M$$

Where $^{(j)}y_i^k$ is the j^{th} perturbed observation for i^{th} state at time point k and $^{(j)}\varepsilon_i^k$ is the amount of perturbation sampled from Gaussian distribution of zero mean and σ_i^k variance in the j^{th} sample. N is the number of states, S is the number of samples and M stands for the number of time points. Variance for perturbation σ_i^k is chosen roughly close to variance of the observation error in i^{th} state and k^{th} time point. (μ_{ik}, see equation (20)). The amount of perturbation , $^{(j)}\varepsilon_i^k$, is selected through Monte Carlo Sampling procedure.

Finally, all data points for each state i at each sample j are collected in a vector form. This is performed for both sides of equation (2). Let,

$$^{(j)}y_i = \left[{}^{(j)}y_i^1,\ldots,{}^{(j)}y_i^k,\ldots,{}^{(j)}y_i^{M-1}\right]^T$$

$$^{(j)}y_i' = \left[{}^{(j)}y_i^2,\ldots,{}^{(j)}y_i^k,\ldots,{}^{(j)}y_i^M\right]^T \tag{16}$$

Equation (16) can be written in a matrix form for all states, $i = 1,\ldots,N$

$$Y^{(j)} = \left[{}^{(j)}y_1 \ \ldots \ {}^{(j)}y_i \ \ldots \ {}^{(j)}y_N\right]^T$$

$$Y'^{(j)} = \left[{}^{(j)}y_1' \ \ldots \ {}^{(j)}y_i' \ \ldots \ {}^{(j)}y_N'\right]^T \tag{17}$$

Next, equation (17) can be written for all samples, $j = 1,\ldots,S$;

$$Y^{Total} = \left[Y^{(1)} \ldots \ Y^{(j)} \ldots \ Y^{(S)}\right]$$

$$Y'^{Total} = \left[Y'^{(1)} \ldots \ Y'^{(j)} \ldots Y'^{(S)}\right] \tag{18}$$

The parameters are estimated using least squares solution;

$$A^T = \left(\left(Y^{Total}\right)^T \cdot Y^{Total}\right)^{-1} \left(Y^{Total}\right)^T Y'^{\,Total} \tag{19}$$

3. ASSESSING THE PERFORMANCE

To test the performance of our method against least squares (LS), Total Least Squares , and Constrained TLS solutions, we created ensemble of 50 linear time discrete systems with different parameters, each consisting of 10 states. This is achieved by creating random $A^{10\times10}$ matrices. We assumed sparse structure for each $A^{10\times10}$ matrix, therefore the number of non-zero elements are fixed to 30 out of 100 total connections. Each system is simulated for certain number of time points (M) starting from a random initial condition. However, we assumed limited number of data for the systems, as most of the real systems have small number of data (Most biological systems, gene networks, etc.). Multiplicative and additive noise terms are added to the simulation results as follows;

$$\hat{x}_{ik} = x_{ik} + x_{ik}\mu_{ik} + \eta_{ik}$$

$$\mu_{ik} \approx N(0, \sigma_1) \qquad \eta_{ik} \approx N(0, \sigma_2) \tag{20}$$

In this equation, $\hat{x}_{ik}$ is the observed value of i^{th} state at k^{th} time point. ε_{ik} and η_{ik} are random variables assumed to have Gaussian distribution with zero mean and variances σ_1 and σ_2 respectively. The term, $x_{ik}\mu_{ik}$ corresponds to the multiplicative noise term, whereas η_{ik} stands for the additive noise term. Our algorithm is tested against the other algorithms for different level of noise and number of time points. Performances of methods for parameter estimation are quantified as the Frobenius norm of the deviation of estimated parameters from their true values relative to the Frobenius norm of true parameters according the following formula;

$$E_A = \frac{\left\|A - A^R\right\|_F}{\left\|A^R\right\|_F} \tag{21}$$

where A^R and A stand for the true and estimated parameters respectively. In addition, fitness of the system is evaluated and compared to true values of states similar to (15);

$$E_X = \frac{\left\|X - X^R\right\|_F}{\left\|X^R\right\|_F} \tag{22}$$

where X and X^R indicates estimated and true values of states, respectively. E_A and E_X are calculated for ensemble of 50 different systems at each number of sample point and averaged. In figure 1, one can see the comparison of the methods with respect to number of samples when multiplicative and additive noise terms are set to $\sigma_1 = 0.10$, and $\sigma_2 = 0.0001$ and number of time points is 10.

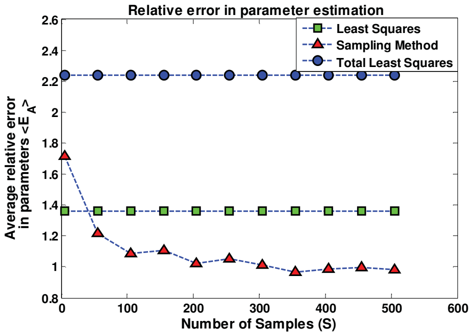

Fig.1. Relative error in parameter estimation vs number of samples for 10 time points at $\sigma_1 = 0.10$

Our method shows significant decrease in relative error in parameters compared to total least squares and least squares with increasing number of samples.

Methods	M=12	M=18	M=24
Least Squares	2730.1	1.1925	0.59776
Sampling Method **(at S=500)**	**0.11563**	**0.48125**	**0.44043**
Total Least squares	200.95	29.465	35.352
Constrained TLS	0.72001	289.98	3.7678

Table1. Average relative error in fitness for different methods for different time points at $\sigma_1 = 0.10$

Table 1 depicts the average relative error in fitness across the different methods and for different number of data. Our method outperforms all methods for 500 samples.

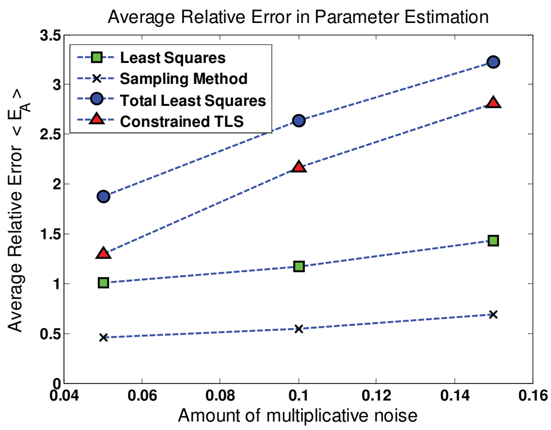

Fig.2. Average relative error in parameters versus amount of multiplicative noise for 18 time points

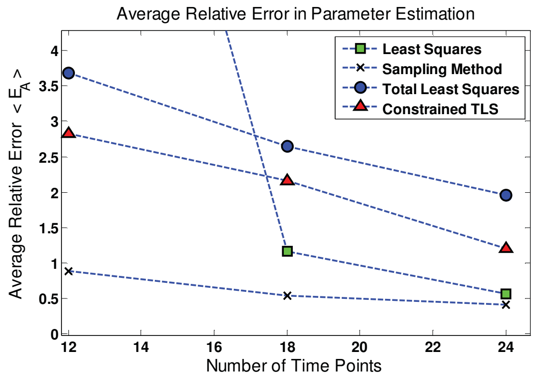

Fig.3. Average relative error in parameters versus number of time points at $\sigma_1 = 0.10$

Figure 2 indicates that our method performs significantly better than LS, TLS and constrained TLS across different levels of noise. TLS is expected to perform better in the case of noisy dependent and independent variables if enough data points are available. However, in this particular problem, due to the serially correlated multiplicative error and limited number of data, its performance is even below least squares solution. We assumed relatively small number of data, because for most of the interesting real systems, data is usually limited and noisy. Constrained TLS solution resolves the serial correlation problem and performs better than TLS, however, its performance still falls behind least squares solution because of limited data. When the number of data increases, the performances of all methods converge (See Figure (3)).

In Fig.3, one can observe that sampling method gives least amount of error in parameters at different number of time steps.

Methods	$\sigma_1 = 0.05$ $\sigma_2 = 0.0001$	$\sigma_1 = 0.10$ $\sigma_2 = 0.0001$	$\sigma_1 = 0.15$ $\sigma_2 = 0.0001$
Least Squares	7.3e+007	2730.1	9.9e+10
Sampling Method **(at S=500)**	**0.059484**	**0.11563**	**0.16913**
Total Least squares	11.331	200.95	21.864
Constrained TLS	1.2951	0.72001	8.4658

Table 2. Average relative error in fitness for different methods for 10 time points at different levels of noises.

In table 2, it is seen that the fitness of our method is much better than other methods for different noise levels.

4. CONCLUSIONS

The contribution of this work can be summarized in two ways. First, our method outperforms the common estimation methods in parameter estimation in the presence of correlated noise. Second, fitness of the estimated parameters through our method is significantly better than LS, TLS and Constrained TLS method. This method is particularly promising in the application of gene network identification problem. The biological measurements are notorious for having a high level of multiplicative noise which makes the network identification problem difficult. Our method has the potential to be the valuable tool for this difficult problem.

REFERENCES

Cadzow, J. A., Wilkes, D. M. (1985). *IEE international conference on Acoustics, Speech and Signal Processing,* 10, 1341-1344.

Huffel, S. V. (1991). *The total least squares problem: computational aspects and analysis,* Society for Industrial and Applied Mathematics, Philadelphia.

Kim, J., Bates, D. G., Postlethwaite, I., Harrison, P., and Cho, K. (2007). *BMC Bionformatics,* 8, 8.

Ljung, L. (1987). *System identification: theory for the user: computational aspects and analysis,* Prentice Hall information and system science series, Englewood Cliff, NJ.