PARAMETER ESTIMATION IN ORDINARY DIFFERENTIAL EQUATIONS

A Thesis

by

CHEE LOONG NG

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

May 2003

Major Subject: Computer Science

PARAMETER ESTIMATION IN ORDINARY DIFFERENTIAL EQUATIONS

A Thesis

by

CHEE LOONG NG

Submitted to Texas A&M University
in partial fulfillment of the requirements
for the degree of

MASTER OF SCIENCE

Approved as to style and content by:

Bart Childs (Chair of Committee)	Daren B.H. Cline (Member)
Vivek Sarin (Member)	Valerie E. Taylor (Head of Department)

May 2003

Major Subject: Computer Science

ABSTRACT

Parameter Estimation in Ordinary Differential Equations. (May 2003)

Chee Loong Ng, B.S., National University of Singapore

Chair of Advisory Committee: Dr. Bart Childs

The parameter estimation problem or the inverse problem of ordinary differential equations is prevalent in many process models in chemistry, molecular biology, control system design and many other engineering applications. It concerns the reconstruction of auxillary parameters by fitting the solution from the system of ordinary differential equations(from a known mathematical model) to that of measured data obtained from observing the solution trajectory.

Some of the traditional techniques (for example, initial value technques, multiple shooting, etc.) used to solve this class of problem have been discussed. A new algorithm, motivated by algorithms proposed by Childs and Osborne(1996) and Z. F. Li's dissertation(2000), has been proposed. The new algorithm inherited the advantages exhibited in the above-mentioned algorithms and, most importantly, the parameters can be transformed to a form that are convenient and suitable for computation. A statistical analysis has also been developed and applied to examples. The statistical analysis yields indications of the tolerance of the estimates and consistency of the observations used.

To Laura, my family, and my bosses in Temasek Polytechnic

ACKNOWLEDGMENTS

My greatest gratification and appreciation goes to Professor Childs, Professor Cline and Professor Sarin for their guidance and assistance. They have made my learning experiences at Texas A&M a fruitful and memorable one. In particular, I would like to specially thank Dr. Childs for his constant encouragement, patience, and confidence in me. He has been instrumental in guiding me in this long, but satisfying, learning journey. I am also greatly motivated by his passion for learning and teaching. This is something that I will carry with me in my teaching profession. My gratification also go to all the computer science faculty members, who have guided me and taught me at Texas A&M.

My dream to study in U.S. would not have been possible if not for the opportunity given to me by Temasek Polytechnic, in particular, my bosses (Mrs. Lay, John Leong and May Lin). I would like to thank them for having confidence in me.

Lastly, I would like to thank my family for standing behind me all these years. I am deeply indebted to them and grateful for what they have given me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

A. Regression Analysis

The technique of regression is a useful statistical tools for the exploration of one variable on others. There are different forms of regression analysis, namly simple, parametric, non-parameter, etc. In this section, I shall focus only on simple linear regression as it is needed in our discussion.

I will use a definition of regression analysis as being fitting a model to data.

In simple linear regression, we usually have a relationship of the form

$$Y_i = \alpha + \beta x_i + \epsilon_i \tag{1.1}$$

where Y_i is a random variable and x_i is another observable variable. It is common to assume that $E(\epsilon_i) = 0$. A major purpose of regression analysis is to predict Y_i from knowledge of x_i. Y_i is usually referred to as the dependent variable and x_i, the independent variable. Using statistical notation, we can state our inferences as

$$E(Y_i|x_i) = \alpha + \beta x_i \tag{1.2}$$

According to Casella & Berger [1], when we perform a regression analysis we investigate the relationship between a predictor and a response variable. There are two steps to the analysis. The first step is a data-oriented one, in which we attempt only to summarize the observed data. We do not make any assumptions about parameters as we are interested only in the data at hand. The second step in the regression analysis is the statistical one, in which we attempt to infer conclusions about the

---

The journal model is *IEEE Transactions on Automatic Control.*

relationship in the population. To do this, we need to make assumptions about the population.

I will be concerned with a particular problem that meets the above definition of regression analysis. This problem is the estimation of unknown parameters in differential equations where I have observations of the response (solution) of the differential equation. Further, these observations will be noisy (due to noise or limited precision). The number of these will be significantly greater than the number required to uniquely determine a solution. This is often called parameter estimation (PE) [2].

We are usually not concerned with the second step of regression analysis when doing PE. This is because in the PE problem we usually have significant detailed knowledge of the ODE model as it is based on concrete theoretical and extensive scientific analysis.

The linear relationship between the independent variables and the observations in the PE problem arises due to the application of Newton's linearization technique [3] on the differential equation and the finite difference approximation on the derivatives. As such, the assumption in 1.2 may not be fully justified. We do, however hope that

$$E(Y_i|x_i) \approx \alpha + \beta x_i \tag{1.3}$$

will be a reasonable approximation.

B. The Boundary Values Problem in Ordinary Differentiation Equations

The general interest to the linear multi-points boundary values problem in ODE is to find the solution to the differential equation subject to m boundary conditions [4]. This problem can be formulated mathematically as follows:

$$\dot{y} = g(y, x) \tag{1.4}$$

subject to the m boundary conditions:

$$q(y(x_i)) = b_i \tag{1.5}$$

where:

y is the state vector of n elements,

x is the independent variable, often *time.*

q_i is an operator that defines a linear combination of the elements of the state vector, y, that is equal to the boundary value b_i at $x = x_i$.

The boundary values conditions are considered at many values of the independent variables.

These boundary value problems can be solved by use of:

1. superposition methods which are also known as shooting methods. Childs' codes have used these [5].

2. finite difference methods.

3. spline or finite element methods.

The last two should enable faster solutions and also has common applicability in partial differential equations problems.

We are concentrating on the second because the last one has a function minimization that conflicts with the use of the least squares process for the specification of the observations.

C. Analysis of Variance

Analysis Of Variance (ANOVA) is a commonly used tool to identify sources of variability from two or more independent potential sources. It is concerned with analyzing variation of means. It provides a useful way of thinking about the way in which different treatments affect a measured variable — the idea of allocating variation to different sources. The ANOVA procedure makes the following assumptions:

- Independence of samples;
- Normality of sampling distributions;
- Equal variance of groups;

Consider the terms in the one-way ANOVA model

$$Y_{ij} = \theta_i + \epsilon_{ij}$$

where the θ_i are unknown parameters, ϵ_{ij} are error random variables, $i = 1, \ldots, k$, and $j = 1, \ldots, n$. Under these assumptions, in particular if $Y_{ij} \sim n(\theta_i, \sigma^2)$, it can be shown that

$$(1/\sigma^2) \sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y_i})^2 \sim \chi^2_{N-k} \tag{1.6}$$

where $\overline{Y_i} = (1/n_i) \sum_j Y_{ij}$ [1]. Furthermore, if the parameters of concerned are the same for every i and j, then

$$(1/\sigma^2) \sum_{i=1}^{k} n_i (\overline{Y_i} - \overline{\overline{Y}})^2 \sim \chi^2_{k-1} \tag{1.7}$$

and

$$(1/\sigma^2) \sum_{i=1}^{k} (\overline{Y_i} - \overline{\overline{Y}})^2 \sim \chi^2_{N-1} \tag{1.8}$$

where $\overline{\overline{Y}} = \sum_i n_i \overline{Y_i} / \sum_i n_i$

Table I. ANOVA table

Source	Degrees freedom	Sum of squares	Mean square	F statistic
Treatment	k-1	SSB=$\sum n_i(\overline{y_i} - \overline{\overline{Y}})^2$	MSB=SSB/(k-1)	F=MSB/MSW
Error	N-k	SSW=$\sum\sum(y_{ij} - \overline{y_i})^2$	MSW = SSW/(N-k)	
Total	N-1	SST=$\sum\sum(y_{ij} - \overline{\overline{Y}})^2$		

It is common to summarize the results of an ANOVA F test in a standard form called an ANOVA table as in Table I adapted from [1].

D. Method of Maximum Likelihood Estimation

The method of Maximum Likelihood Estimation(MLE) is by far, the most popular techniques for deriving estimators. Suppose if $X_1, \cdots, X_n$ are an independent and identically distributed random variables from a population with pdf or pmf $f(x|\theta_1, \cdots, \theta_k)$, the likelihood function is defined by

$$L(\theta|\mathbf{x}) = L(\theta_1, \cdots, \theta_k|x_1, \cdots, x_n) = \prod_{i=1}^{n} f(x_i|\theta_1, \cdots, \theta_k) \tag{1.9}$$

A maximum likelihood estimator(MLE) of the parameter θ based on a sample $\mathbf{X}$ is $\hat{\theta}(\mathbf{X})$ where $\hat{\theta}(\mathbf{x})$ is a parameter value at which $L(\theta|\mathbf{x})$ attains its maximum as a function of θ, with $\mathbf{x}$ being held fixed.

It is interesting to note that the least square estimators of α and β in equation 1.1 are also the MLE of α and β [1].

E. Parameter Estimation Problem

Consider ordinary differentiation equation of the following form:

$$\dot{y} = f(x, y, \theta) \tag{1.10}$$

subject to the following boundary conditions:

$$bc_i(x_i, y_i, \theta) = 0 \tag{1.11}$$

where $x \in [a, b]$ the independent variable, $y = y(x, \theta)$ denotes the state vector depending on x and θ ($\in \Re$) denotes the parameter vector and lastly bc_i denotes the boundary condition constraints observed at x_i.

Let m be the number of measurements available from the process, $\hat{y}_i$ denotes the i^{th} observed values (termed *observation*), given in terms of functions of states and parameters at data points x_i and ϵ_i denotes the inherent errors presence in this measurement.

The primary goal in parameter estimation problem is to identify reasonable values for the parameter θ by fitting a theoretical model to the measured data points.

In this model, we assume that ϵ_i is an independent and identically distributed random variable which follows a normal distribution with mean 0 and variance σ^2.

We aim to minimize

$$\sum_{i=1}^{i=m} \epsilon_i^2 \tag{1.12}$$

subject to the least square constraints. The method of MLE and regression analysis are commonly used statistical tools to identify the parameters and perform the required analysis.

The available data may arise from different processes under different conditions and the parameter estimation problem thus consists of several independent problem sets which have only the parameters θ in common [6]. This is where we can apply the idea of ANOVA to perform the required analysis. In this instance, we aim to analyse the variation in the mean between the data points obtained from these different sets.

CHAPTER II

HISTORY

The parameter estimation problem is also called by the names: system identification problem, parameter identification problem, and inverse problem [2] & [4]. The estimation of parameters — or unknown initial values — given data on the trajectory of a differential equation system. It involves the fitting of the solutions of a system of differential equations to data corresponding to a realization of a particular solution trajectory observed in the presence of noise [7]. We are particularly concerned with estimating parameters in the case of a non-linear multi-point boundary values ordinary differential equations (ODE) because of application potential in many process models [2].

Typically, the approach to this problem is solved by *initial valve problem* (IVP) methods [4], [8] & [9]. It involves the repeated solution of the initial value problem using some forms of iterative algorithm to improve the fit. This approach has several drawbacks, which was discussed by [6]. It was shown by many that this approach causes deterioration of efficiency as it focuses on the parameters and thus neglecting inherent state information on the inverse problem. Another reason is that this elimination of the state information can result in a substantial loss of stability for the solution structure. This is especially pertinent for the case when bad initial parameter values were chosen. In this instance, the IVP may be ill-conditioned and hard to solve. In some instances, even when the parameter problem is perfectly well-conditioned, a solution may not even exist. This situation was illustrated by Bock in [6], which he termed as a "notorious test" problem. In that problem, the ODE for two states and one unknown parameter p with fixed initial values is given by:

$$\dot{x_1} = x_2 \qquad\qquad \dot{x_2} = \mu^2 x_1 - (\mu^2 + p^2) sin(pt) \tag{2.1}$$

where $x_1(0) = 0$, $x_2(0) = \pi$ and $t \in [0, 1]$. The solution of the *true parameter value* p=π is $x_1(t) = sin\pi t$, $x_2(t) = \pi cos\pi t$. It has been shown that for true value of p (correct to 16 decimals) and with the highest integration accuracy, the solution is properly reproduced only on the first half of the interval! [6]. The problem is characterized by having positive and negative eigenvalues.

Many of these problems can be solved using Multiple Shooting Methods [10]. This method involves the superposition of the initial value solutions of the differential equations over short subintervals and enforcing continuity of the solution across the boundaries of the intervals. It involves the conversion of the original second order ODE to two first order equations, parameterization of the parameters and making an initial estimates of the parameters. The systems of differential equation is then integrated over the intervals. This approach results in the solving of a constrained over-determined systems, usually by using the least square constraints approach. A point to note is that a good selection of mesh is neccessary. This is to avoid drifting too far away from the solution trajectory [6].

Once we have decided on a suitable parameterization, we can then proceed with the matching of the computed and observed data points in order to estimate the auxillary parameters. Childs and Osborne [10] have proposed a stable and efficient algorithm in handling the matching between the solutions (of the ODE) and the observed data. Minimization of the sum of squares of the discrepanies between these data points is carried out using Fisher's method of scoring [11]. Fisher's method incorporates the idea of Maximum likelihood (MLE) and Gauss-Newton methods in its approach. It has thus inherited some of the advantages (quadratic rate of convergence and good transformation of invariance properties) associated with Gauss-Newton's method. In addition, it only requires first derivative information. An implementation code, `ps_quasi` has also been provided by Childs in [5].

Li [12] in his dissertation, has proposed a method addressing the issues of explicit parameterization and restrictions (large sample size and the correctness of model formulation) associated with Gauss-Newton's method. The proposed method first transforms the ODEs estimation problem into a non-linear programming problem by applying the finite difference method, where both the state variables and the parameters are regarded as unknown variables. The system of equations to be mimimized can be extensively huge when there is a large sample of observation data. The advantage is that it has few degrees of freedom. Cyclic reduction is then applied to reduce it to a minimization problem with a fixed number of constraints. This does not require explicit imposing of extra initial or boundary conditions. The author has also made use of Sequential Quadratic Programming (SQP) method [12] to relax the restrictions imposed by Gauss-Newton methods.

CHAPTER III

PROPOSED WORK

Li's dissertation has provided an in-depth discussion of the various methods of getting the solutions to the estimation problem. This work can be viewed as an extension to his approach with an appropriate statistical analysis of the results.

A new algorithm has been developed for solving this class of problem, motivated by algorithms proposed by [5] and [12]. This new algorithm is summarized in Fig. 1.

This new algorithm has inherited some of the merits exhibited in the algorithms of [5] and [12]. One of them is the quadratic convergence of Gauss-Newton's method. In addition, the parameterization (of the unknown parameters) can be carried out easily and convergence is rapid in most cases. In the solving of over-determined system of equations, Li made use of SQP techniques. I use the approach of Childs which he had used in his implementation in [5]. To faciliate the statistical analysis, a small ODE solver(written in Matlab Version 6.5) implementing the algorithm in Fig. 1 has been developed. An outline of the implementing codes is in Appendix B.

A. Problem: Spring Mass Dashpot Model

The model of a spring mass dashpot, electrical circuit, or other physical systems is used as a sample problem for the simple solver to solve. This model has been implemented by Childs in [5] and the numerical results have been presented and analyzed in [10]. The results obtained in my thesis will provide a good basis of comparision between the two approaches.

This problem is modelled mathematically as follows:

$$\ddot{x} + \mu\dot{x} + \xi x = \lambda sin(t) \tag{3.1}$$

1. Parameterization of the state vector and the unknown parameters by a linearization about a reference solution, w, on the specified ODE — this is common to Childs and Osborne's work;

2. Application of the finite difference technique to the state vector and its derivatives. This results in a system of algebraic equations rather than ODEs;

3. Forming a system of algebraic equations that is to be solved as a constrained least squares problem;

4. Solving the system of equation using a constrained least squares approach;

5. Iteration of steps 3 through 4 using the newly computed results as the referenced solutions, until the specified iterations or the desired tolerance level is reached;

6. Calculation of the statistics reflecting the quality of the fit.

Fig. 1. Proposed new algorithm

subject to the $n + 1$ boundary conditions:

$$q_i(x(t_i)) = b_i \tag{3.2}$$

for $i = 1, 2, \ldots, n + 1$. Notice that the notorious test problem can be obtained with specific values of the parameters μ, ξ and λ. Childs'codes, using multiple shooting has been used to solve this problem, though unpublished.

The constants μ, ξ and λ are unknown and the estimation of these parameters are of primary focus.

The differential equation is linearized about a reference solution on the parameter and the state variables and finite difference method is then applied to the equation.

This is what we have:

$$[2 + h\mu_0]x_{i+1} + [-4 + 2h^2\xi_0]x_i + [2 - h\mu_0]x_{i-1}$$
$$[h(w_{i+1} - w_{i-1})]\mu + [2w_i h^2]\xi - \lambda 2h^2 sin(t_i) = [w_{i+1} - w_{i-1}]h\mu_0 + 2\xi_0 w_i h^2 \quad (3.3)$$

where w, μ_0 and ξ_0 are reference solutions or assumed initial estimates.

Suppose we have n partitions in the mesh, $n + 1$ observations to be met in the "least square sense" and $n - 1$ equations are "exact" (from the finite difference method). We will then have a system of equations of the following form:

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1(n+1)} & a_{1(n+2)} & a_{1(n+3)} & a_{1(n+4)} \\ 0 & a_{22} & \ddots & \cdots & a_{2(n+2)} & a_{2(n+3)} & a_{2(n+4)} \\ 0 & 0 & \ddots & \cdots & a_{3(n+2)} & a_{3(n+3)} & a_{3(n+4)} \\ 0 & 0 & \ddots & \ddots & \vdots & \vdots & \vdots \\ 0 & \ddots & \ddots & \ddots & a_{(n-1)(n+2)} & a_{(n-1)(n+3)} & a_{(n-1)(n+4)} \\ 1 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 1 & \cdots & \cdots & \cdots & \cdots & \cdots \\ \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \cdots & \cdots & \cdots & 1 & 0 & 0 & 0 \end{bmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_{n+2} \\ x_{n+3} \\ x_{n+4} \end{pmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ \vdots \\ \vdots \\ b_{2n} \end{bmatrix} \quad (3.4)$$

Using block notation, we can partition this large matrix into four block of smaller matrices of the following form:

$$\begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix} \begin{pmatrix} x_e \\ x_l \end{pmatrix} = \begin{bmatrix} b_e \\ b_l \end{bmatrix} \quad (3.5)$$

where the matrices A_1, A_2, A_3, and A_4 are of order $(n - 1) \times (n - 1)$, $(n - 1) \times 5$,

$(n+1) \times (n-1)$, and $(n+1) \times 5$ respectively. x_e and x_l are the respective "exact" and "least square" state vectors. The $(n-1) \times (n-1)$ tri-diagonal matrix, A_1, is of the following form:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & \cdots & 0 \\ 0 & a_{22} & a_{23} & a_{24} & \cdots & 0 \\ 0 & 0 & a_{33} & a_{34} & \cdots & 0 \\ \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \cdots & \cdots & \cdots & \cdots & a_{(n-1)(n-2)} & a_{(n-1)(n-1)} \end{bmatrix}$$

The system of equations

$$\begin{bmatrix} A_1 & A_2 \end{bmatrix} \begin{pmatrix} x_e \\ x_l \end{pmatrix} = \begin{bmatrix} b_e \end{bmatrix} \tag{3.6}$$

is to be solved exactly while computing the least square solution for these over-determined system:

$$\begin{bmatrix} A_3 & A_4 \end{bmatrix} \begin{pmatrix} x_e \\ x_l \end{pmatrix} = \begin{bmatrix} b_l \end{bmatrix} \tag{3.7}$$

This system of equations is then reduced to the following form:

$$\begin{bmatrix} A_1' & A_2' \\ A_3' & A_4' \end{bmatrix} \begin{pmatrix} x_e \\ x_l \end{pmatrix} = \begin{bmatrix} b_e' \\ b_l' \end{bmatrix} \tag{3.8}$$

where A_1' is an identity matrix, A_3', a zero matrix, and b_e' and b_l' are the corresponding vectors on the right hand side of the equation after reduction.

x_l is obtained by solving the $(n+1) \times 5$ systems of equations

$$A_4' * x_l = b_l' \tag{3.9}$$

in a "least square sense" and subsequently x_e is obtained from

$$x_e = b'_e - A'_2 * x_l \tag{3.10}$$

B. Statistical Approach

At this point, I will discuss the statistical approaches and methodologies that will be used. This portion of the discussion is an adaption from [7].

Forming the "normal equation" in 3.7, we have

$$\begin{bmatrix} (A'_4)^T & A'_4 \end{bmatrix} \begin{pmatrix} x_l \end{pmatrix} = \begin{bmatrix} (A'_4)^T * b_l \end{bmatrix} \tag{3.11}$$

We shall denote the normal equations by

$$N * x_l = (A'_4)^T * b_l \tag{3.12}$$

where

$$N = (A'_4)^T * (A'_4) \tag{3.13}$$

Let SSE denotes the sum of squares of errors. The variance of x_l is:

$$V(x_l) = N^{-1} * s^2 \tag{3.14}$$

where s^2=SSE/(degrees of freedom). The degree of freedom is computed by taking the difference between the total number of equations and total number of unknowns. The variance of x_e is:

$$V(x_e) = (A'_2)V(x_e)(A'_2)^T \tag{3.15}$$

The covariance matrix for $\mathbf{x}$ is:

$$V(\mathbf{x}) = \begin{bmatrix} V(x_e) & 0 \\ 0 & V(x_l) \end{bmatrix}. \tag{3.16}$$

The prediction interval of an observation is computed based on the following [13]:

$$\hat{b_i} \pm t(m - (n+4), \alpha/2)\sqrt{s^2 + V(i,i)} \tag{3.17}$$

where $V(i,i)$ is the diagonal element at position (i,i) of $V(\mathbf{x})$. The second term in this equation is an indication of the "tolerance of the estimate" and is a good indicator of quality of the estimation process for the unknown state which includes the solution of the ODE and its parameters. The term *tolerance of the estimate* will be used interchangeably with *confidence estimate* in this thesis.

CHAPTER IV

COMPUTATIONAL RESULTS

A. Test Cases for ODE Solver

The solver was first validated modularly and then verification was then carried out as a whole to determine if the coding was done correctly.

1. Testing the Constrained Least Squares Method

The code of [5] includes a Fortran 90 implementation of the algorithm developed starting with equation 3.5. This is done using a Gauss Jordan maximum pivot strategy with the least squares part using a singular value decomposition strategy. My development was done using MATLAB 6.5 and it is recognized that there are possible problems where the strategy used by Childs may be preferable.

The validity of my constrained least squares was tested by use of intermediate results from Childs' code being furnished for test cases. The differences were small and deemed to be acceptable. These were likely caused by the slight increases in precision available from the use of pivoting and singular value decomposition in Childs' code.

B. Result for Spring Mass Dashpot Problem

I will use a problem that was used by Childs [5] to give me a basis of comparison. The ODE is second order and there are three unknown parameters. I will use observations on the same fixed range of the independent variable. I will divide this into n intervals where n will be 16, 32, 64, 128, and 256.

The *state* of the problem will be $n+4$ from the three parameters and $n+1$ values of the dependent variable of the ODE at end-points of the intervals.

Table II. Summary of specific dependent variables at each iteration for 32 intervals

	Independent variables t					
Iter.\t	0.0	4.0	8.0	12.0	16.0	Norm($w-x$)
0	-3.5889576336	7.5278558980	-3.4424224401	-0.1575283205	5.6609498154	
1	-2.9967033632	6.2783624427	-2.8704563483	-0.1243323378	4.7227238309	4.7747e+000
2	-2.9962886298	6.2781553821	-2.8697644659	-0.1240454630	4.7233446345	4.6044e-002
3	-2.9961913066	6.2781235934	-2.8697877596	-0.1240425389	4.7234108259	3.7575e-004
4	-2.9961912858	6.2781236180	-2.8697877681	-0.1240425108	4.7234107807	1.2330e-007
5	-2.9961912858	6.2781236180	-2.8697877680	-0.1240425109	4.7234107808	1.6414e-010
6	-2.9961912858	6.2781236180	-2.8697877680	-0.1240425109	4.7234107808	3.3702e-013

A random noise has been added to the observations and we have started the iteration with the reference solutions multiplied by a user-specified perturbation (in this case, it was specified as 1.2). The parameters values of μ and ξ are 25% off the true values. The iterative methods stop when we have achieved the specified iterations (maximum of 8 iterations in my program) or when the desired norm of the solutions (norm$(w-x) < 0.5e-10$) has been reached, whichever comes first. The estimates of dependent variables $x(t_i)$ where $t_i = 0, 4, 8, 12, 16$ for the 32 intervals are shown in Table II at each iteration. The estimates of the state vectors and the parameters close in well to the analytical solution and the true values at the first iteration. This has demostrated the rapid convergence of the algorithm. From Tables II & III, we noticed that we have achieved the quadratic convergence. The quadratic convergence was not too obvious towards the later iterations. Detailed results of the estimate, lower limit, upper limit & confidence estimate are as shown in Table IV. A plot of the test solutions is shown in Fig. 2. A scatter plot showing the difference between the numerical solutions & the observation is reported in Fig. 3. One can observe that all the points lie within the limits of the tolerance estimate.

From Table V, we can clearly observe that the values of the parameter values are closer to the true parameter values when the mesh is increased. This is definitely in-line with what we expect from the result. It was not so clear cut in the case of the confidence estimates. The confidence estimates for 32 intervals are larger than that

Table III. Summary of parameter values at each iteration

Iteration	mu	xi	lambda	sqrt(2-norm)
0	0.1500000000	1.2500000000	1.0000000000	
1	0.1922049830	1.0240557389	0.9595342732	2.3339e-001
2	0.2003692255	0.9790424175	0.9578898754	4.5777e-002
3	0.2003658136	0.9790424087	0.9578989751	9.7183e-006
4	0.2003658059	0.9790424176	0.9578989100	6.6137e-008
5	0.2003658059	0.9790424176	0.9578989101	8.3486e-011
6	0.2003658059	0.9790424176	0.9578989101	1.7699e-013

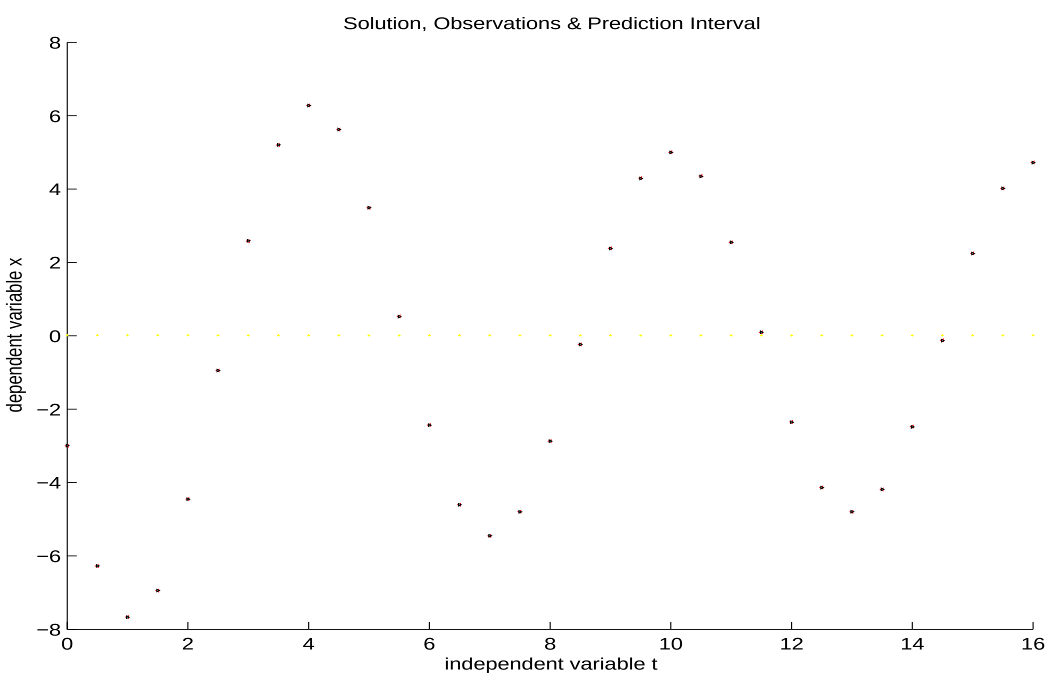

Fig. 2. Plot of test solutions

Table IV. Summary of statistics for 32 intervals

Specifics for the boundary conditions & parameter estimates

t_alpha(alpha=0.05)=2.05 d.f=28

t	estimate	lower limit	$\leq$	observation	$\leq$	upper limit	est. tol
0.00	-2.996191	-3.010333	$\leq$	-2.990798	$\leq$	-2.982050	0.014142
0.50	-6.272323	-6.286241	$\leq$	-6.270183	$\leq$	-6.258405	0.013918
1.00	-7.664581	-7.678363	$\leq$	-7.667390	$\leq$	-7.650798	0.013783
1.50	-6.945614	-6.959233	$\leq$	-6.940782	$\leq$	-6.931994	0.013620
2.00	-4.448840	-4.462219	$\leq$	-4.452971	$\leq$	-4.435462	0.013378
2.50	-0.945945	-0.959023	$\leq$	-0.946393	$\leq$	-0.932867	0.013078
3.00	2.579729	2.566873	$\leq$	2.589216	$\leq$	2.592585	0.012856
3.50	5.199926	5.187107	$\leq$	5.204425	$\leq$	5.212744	0.012818
4.00	6.278124	6.265247	$\leq$	6.273213	$\leq$	6.291000	0.012877
4.50	5.617531	5.604686	$\leq$	5.625127	$\leq$	5.630375	0.012845
5.00	3.487672	3.474997	$\leq$	3.490940	$\leq$	3.500346	0.012675
5.50	0.529402	0.516878	$\leq$	0.522523	$\leq$	0.541926	0.012524
6.00	-2.430933	-2.443494	$\leq$	-2.436697	$\leq$	-2.418371	0.012561
6.50	-4.605946	-4.618696	$\leq$	-4.604082	$\leq$	-4.593197	0.012749
7.00	-5.450819	-5.463710	$\leq$	-5.448769	$\leq$	-5.437929	0.012891
7.50	-4.794757	-4.807612	$\leq$	-4.797949	$\leq$	-4.781902	0.012855
8.00	-2.869788	-2.882482	$\leq$	-2.868685	$\leq$	-2.857094	0.012694
8.50	-0.233939	-0.246496	$\leq$	-0.234915	$\leq$	-0.221382	0.012557
9.00	2.387063	2.374536	$\leq$	2.378382	$\leq$	2.399590	0.012527
9.50	4.295607	4.283036	$\leq$	4.286892	$\leq$	4.308177	0.012571
10.0	5.003687	4.991064	$\leq$	5.000674	$\leq$	5.016311	0.012624
10.5	4.353865	4.341202	$\leq$	4.344466	$\leq$	4.366528	0.012663
11.0	2.550602	2.537915	$\leq$	2.547924	$\leq$	2.563289	0.012687
11.5	0.096821	0.084137	$\leq$	0.099315	$\leq$	0.109505	0.012684
12.0	-2.345075	-2.357740	$\leq$	-2.355021	$\leq$	-2.332410	0.012665
12.5	-4.129769	-4.142473	$\leq$	-4.141046	$\leq$	-4.117065	0.012704
13.0	-4.796733	-4.809598	$\leq$	-4.796026	$\leq$	-4.783868	0.012865
13.5	-4.186200	-4.199276	$\leq$	-4.184513	$\leq$	-4.173124	0.013076
14.0	-2.474870	-2.488013	$\leq$	-2.483583	$\leq$	-2.461726	0.013143
14.5	-0.124043	-0.137020	$\leq$	-0.131274	$\leq$	-0.111065	0.012977
15.0	2.244623	2.231824	$\leq$	2.241902	$\leq$	2.257422	0.012799
15.5	4.012419	3.999395	$\leq$	4.017851	$\leq$	4.025443	0.013024
16.0	4.723411	4.709657	$\leq$	4.717458	$\leq$	4.737164	0.013754

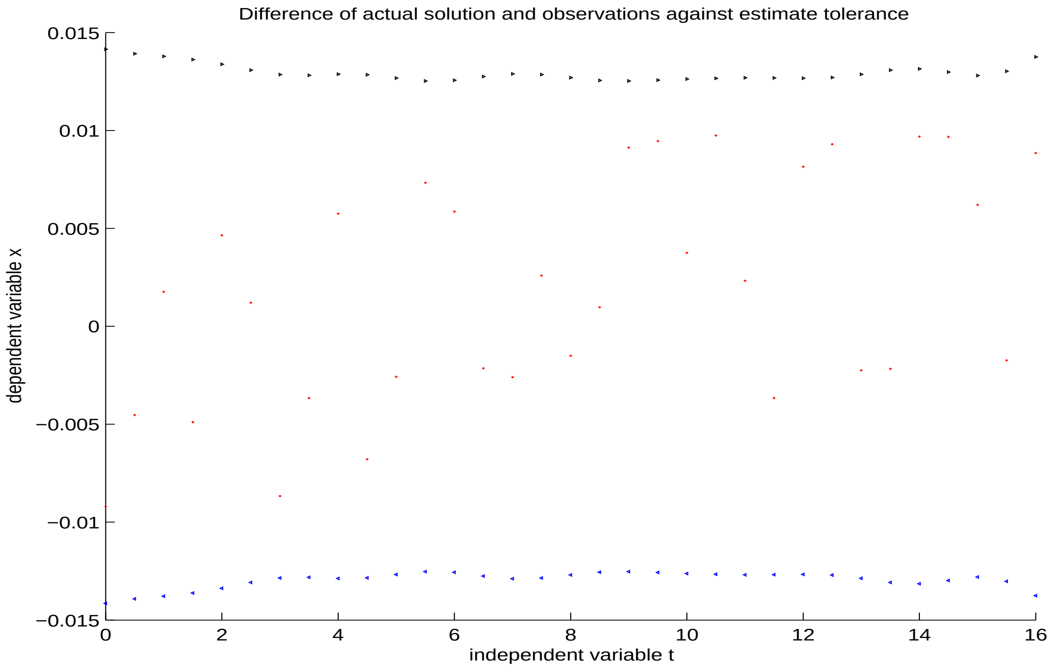

Fig. 3. Observation error against estimates tolerance

of 64 and 128 interval. This is somewhat expected since we have lesser observations. But this was not the case for 64 intervals and 128 intervals. In this instance, it was much smaller in the 64 intervals case. The probable cause could be due to the finite difference approximations.

The parameter values obtained for λ using 16 intervals were off by about 15%. In the case of 256 intervals, the results obtained were very close with the 128 intervals. Both converge in 5 iterations. In fact, better estimates are obtained for μ in the case when we have 128 intervals. The percentage of error for μ is 0.019%(in 128 intervals) as compared to 0.22%(in 256 intervals). The value for λ is somewhat similar at approximately 0.3% off the true value. The estimated value obtained for ξ differs by only about 0.05%. Detailed output for 16 intervals and 256 intervals are as attached in appendix A.

Table V. Comparison of parameters estimate and tolerance estimate for varying numbers of intervals

	No. of Interval(n)				
	16	32	64	128	256
estimate-μ	0.202412	0.200366	0.199636	0.199962	0.200441
estimate-ξ	0.917702	0.979042	0.995276	0.998709	0.999153
estimate-λ	0.845241	0.957899	0.984949	0.997118	1.003011
est. tol.-μ	0.015778	0.012131	0.011433	0.010614	0.011348
est. tol.-ξ	0.015737	0.012129	0.011435	0.010615	0.011348
est. tol.-λ	0.020265	0.014670	0.012841	0.011319	0.011738
s^2	0.000052	0.000035	0.000033	0.000029	0.000033

C. Comparison between Childs's Approach and Proposed Approach

The observations generated from my approach were used with Childs' code [5] to generate estimates for $x(t_0)$, μ, ξ & λ. Results are summarized in Table VI.

A similar pattern was observed between the two approaches. The computed percentage got better as we increased the number of intervals in most cases. This may be due to the characteristics of the noise added. Extensive tests should enable resolution of these anomalies. It may be a result of the finite difference approximation that we have applied.

Table VI. Comparison of Childs's approach & proposed approach

	Childs's approach			*Proposed approach*		
			n=16			
	estimate	confidence estimate	as a %	estimate	confidence estimate	as a %
$x(t_0)$	-2.995	0.1822e-01	0.61	-2.989	0.1987e-01	0.66
μ	0.1995	0.4710e-02	2.36	0.2024	0.1578e-01	7.79
ξ	1.0	0.4709e-02	0.47	0.9177	0.1574e-01	1.71
λ	0.9979	0.2217e-01	2.22	0.8452	0.2027e-01	2.40
			n=32			
$x(t_0)$	-2.992	0.1463e-01	0.49	-2.996	0.1414e-01	0.47
μ	0.1997	0.4387e-02	2.20	0.2004	0.1213e-01	6.05
ξ	1.001	0.4414e-02	0.44	0.9790	0.1213e-01	1.24
λ	0.9978	0.1702e-01	1.71	0.9579	0.1467e-01	1.53
			n=64			
$x(t_0)$	-3.001	0.1070e-01	0.36	-2.999	0.1253e-01	0.42
μ	0.2003	0.3959e-02	1.98	0.1996	0.1143e-01	5.72
ξ	0.9998	0.3980e-02	0.40	0.9953	0.1144e-01	1.15
λ	1.002	0.1212e-01	1.21	0.9849	0.1284e-01	1.30
			n=128			
$x(t_0)$	-2.999	0.8906e-02	0.30	-3.000	0.1116e-01	0.37
μ	0.1999	0.4172e-02	2.09	0.1999	0.1061e-01	5.31
ξ	1.0	0.4185e-02	0.42	0.9987	0.1062e-01	1.06
λ	0.9995	0.9891e-02	0.99	0.9971	0.1132e-01	1.14

CHAPTER V

SUMMARY

I have studied the parameter estimation problem in Ordinary Differentiation Equation (ODE). These are common in many process models in the field of Engineering. This class of problem concerns the re-construction of auxillary parameters by fitting the numerical solutions to that of the measured data from observing the solution trajectory.

I have briefly discussed some of the commonly used techniques for parameter estimation problem in Chapter I. In particular, Linear Regression, Maximum Likelihood Estimators (MLE), ANOVA approaches were discussed. The traditional ANOVA is not applicable to this problem and a similar display based on confidence estimates is shown.

I went on to relate those techniques that are applicable to the parameter estimation problem in ODE and how we can adapt it to the solving of this class of problem. Linearity between the observations and the independent vectors is due to the application of Newton's linearization and the finite difference approximation. Some of the traditional techniques used in the solving of ODE were discussed in Chapter II.

I have used algorithms from [5] and [12]. In Chapter III, the main focus of my thesis is introduced. I have proposed a new algorithm combining the merits seen in [5] and [12]. I discussed and explained the statistical approach used in the analysis. To faciliate the statistical analysis process, I have developed a small ODE solver using Matlab 6.5. I have applied to the well-known spring mass dashpot problem and results obtained and some findings were presented in Chapter IV. Some thoughts on future research areas are proposed in Chapter VI.

CHAPTER VI

FUTURE WORK

This thesis has provided some statistical analysis of the fit of numerical solutions of ODEs to simulated observations. This is particularly helpful in aiding the understanding of the relationship between the observations and the parameters in the ODE. This has furthered understanding on the behavior of the observations in relation to its actual analytical solution and the independent vector. Several other aspects and variations of what have been done are worthy of future research and investigations.

1. In [7], the steady state problem was mentioned, which arises as $t \longrightarrow \infty$. It may be interesting to study the behavior of the observations and the parameter values in relation to its independent vectors.

2. In my thesis, I have considered the case where we have $n-1$ exact equations. How will it affect the parameter values if we know the true values of some of the observations? We could vary the number of exact equations and analyse its impact on the parameter values.

3. My examples were based on ODEs. Most engineering and science problems are based on partial differential equations. The statistical procedures developed here are applicable for these problems as well.

4. This methodology can be used by engineers and scientists in many areas. These codes should allow the testing of the constructed device to ensure that the parameters "designed" are reflected in the performance of the actual device.

REFERENCES

[1] G. Casella and R. L. Berger, *Statistical Inference*, 2nd edition, Pacific Grove, CA: Duxbury/Thomson Learning, 2002.

[2] *Numerical Treatment of Inverse Problems in Differential and Integral Equations: Proceedings of an International Workshop, Heidelberg, Fed. Rep. of Germany*, August 30-September 3, 1982. P. Deuflhard, E. Hairer, Eds., Boston: Birkhäuser, 1983.

[3] L. Fox, *The Numerical Solution of Two-Point Boundary Problems in Ordinary Differential Equations*, Oxford: Clarendon Press, 1957.

[4] *Codes for Boundary-Value Problems in Ordinary Differential Equations: Proceedings of a Working Conference*, May 14-17, 1978. B. Childs, M. Scott, J. W. Daniel, E. Denman, P. Nelson, Eds., Berlin, New York: Springer-Verlag, 1979.

[5] B. Childs, " PS_Quasi, an ODE BVP solver," personal collection, 1998.

[6] H. G. Bock, "Recent advances in parameter identification techniques for O.D.E.," in *Numerical Treatment of Inverse Problems in Differential and Integral Equations*, P. Deuflhard, E. Hairer, Eds., Boston: Birkhäuser, 1982, chap. 7, pp. 95–112.

[7] B. Childs, "Some inverse, ill-poised, and/or ill-conditioned problems in ordinary differential equations," paper presented at Australian Mathematical Society - Annual Meeting, Melbourne, July 15-17, 1991.

[8] W. Cheney and D. Kincaid, *Numerical Mathematics & Computing*, 3rd edition, Pacific Grove, CA: Brooks/Cole, 1994.

[9] W. Kaplan, *Ordinary Differential Equations*, Reading, MA: Addison-Wesley, 1958.

[10] B. Childs and M. Osborne, "Fitting solutions of ordinary differential equations to observed data," in *Computational Techniques and Applications: CTAC'95*. R.L. May, A. K. Easton, Eds., Singapore: World Scientific Publishing Co, 1996, pp. 193–198.

[11] M. Osborne, "Fisher's method of scoring," *International Statistical Review*, vol. 60, pp. 99-117, 1992.

[12] Z. F. Li, "Parameter estimations of ordinary differential equations," Ph.D. dissertation, Australia National University, Canberra, 2000.

[13] W. Mendenhall, R. L. Scheaffer and D. D. Wackerly, *Mathematical Statistics with Applications*: Boston, MA: Duxbury Press, 1981.

APPENDIX A

OUTPUT

This appendix contains output obtained for 16 and 256 intervals.

Output for 16 intervals

noise:1.0000e-002

perturbation:1.2000e+000

Iteration	μ	xi	λ	sqrt(2-norm)
0	0.1500000000	1.2500000000	1.0000000000	
1	0.1926216893	0.9736350572	0.8389008659	3.2272e-001
2	0.2024087790	0.9176924375	0.8453444390	5.7157e-002
3	0.2024122387	0.9177021467	0.8452399016	1.0504e-004
4	0.2024123693	0.9177020255	0.8452408038	9.1967e-007
5	0.2024123696	0.9177020252	0.8452408060	2.2603e-009
6	0.2024123697	0.9177020252	0.8452408060	9.7552e-012

Sqrt of 2-norm$(w - x)$

Iteration	dependent variable	parameter	overall
1	3.4012e+000	3.2272e-001	3.4165e+000
2	8.3922e-003	5.7157e-002	5.7769e-002
3	1.0053e-003	1.0504e-004	1.0108e-003
4	1.1658e-006	9.1967e-007	1.4849e-006
5	3.6539e-009	2.2603e-009	4.2965e-009
6	1.3590e-011	9.7552e-012	1.6729e-011

Output for 256 intervals

noise:1.0000e-002

perturbation:1.2000e+000

Iteration	μ	xi	λ	sqrt(2-norm)
0	0.1500000000	1.2500000000	1.0000000000	
1	0.1920047790	1.0409758036	1.0028120232	2.1322e-001
2	0.2004417979	0.9991504054	1.0030275202	4.2668e-002
3	0.2004413589	0.9991525042	1.0030107253	1.6931e-005
4	0.2004413604	0.9991525024	1.0030107385	1.3332e-008
5	0.2004413604	0.9991525024	1.0030107385	1.7813e-012

	Sqrt of 2-norm$(w-x)$		
Iteration	dependent variable	parameter	overall
1	1.3352e+001	2.1322e-001	1.3354e+001
2	3.1863e-003	4.2668e-002	4.2787e-002
3	2.2489e-004	1.6931e-005	2.2553e-004
4	5.1245e-008	1.3332e-008	5.2951e-008
5	1.3666e-011	1.7813e-012	1.3781e-011

APPENDIX B

SOURCE CODES

The simple ODE solver is written in Matlab 6.5. Readers who are interested in looking at the source codes for further research works can request a copy of it from Professor Bart Childs at bart@cs.tamu.edu. I have provided some write-up of each function as follows:

Application Script

This is a simple script that solve ODE of the form:
$x'' + \mu x' + \xi x = \lambda sin(\omega t)$ where μ, ξ and λ are unknown parameters to be estimated. ω is a known constant in the equation. It is set to 1.0 so that we have a period of 2π. The true parameters values are 0.2, 1.0 and 1.0 respectively. When there is a change in the number of data points, or equations, the following may need to be changed:
a - the starting independent variable of the region of interval under consideration;
c - the ending point of the region of intervals;
n - the no. of interval (in the region);
the input(data) file (if any)
p - to indicate no. of exact parameter;
no_exact_eqn - indicate no. of exact equations;
noise - add random noise to the observations;
perturb - perturbation is the deviation of the reference solutions from the observations;
The pseudo-code of the script **test_v4_original_latest.m** is as follows:

```
  script <test_v4_original_latest.m>
%Specification of input values (as mentioned above)
%Calculate the analytic solutions, observations & reference solution w\\
call <analytic_soln.m>
  while(iter <9 & norm(w-x)<0.5e-10)
%Form the system of exact & least squares equations
call <form_eqn3_v4_original_latest.m>
%Perform reduction to the system of equation
call <reduced_matrix_v4.m>
%Solve the constraint least squares equations\\
  call <l_sqr_solver_v4.m>
assign w=x
  end while
%Output
call script <print_output.m>
%Computing the statistics
call <test_stat.m>
```

Analytic Solution

This function is used to compute the exact analytic function. The variables and parameters needed for the analytic solution are:
t: array of independent variable;
n: number of intervals;
noise: to add noise to observations;
perturb: to start the referenced solution at a user-specified perturbation level;

This function will return **analytical solutions, observations & referenced solutions**

Form Equation

This function **form_eqn3_v4_original_latest.m** is used to form the system of algebraic equations after the application of the finite difference techniques and Gauss-Newton linearization.
The input to this function are *array of independent variables*, *number of intervals*, *the gap between independent variables*, *array of observations*, *array of referenced solutions* and lastly, *a flag array* to indicate whether the parameters exact or not.

This function will return a matrix A of order $2n \times (n+4)$ and the corresponding column vector b from the system of algebraic equations $Ax = b$.

Reduced Matrix

This function **reduced_matrix_v4.m** is used to convert the matrix to a form that is suitable for us to apply the least square constraint method. The input to this function are the matrix A, column vector b in the system of algebraic equation $Ax = b$.

This function will return a matrix B of the form: $\begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix}$ where A_1 is a $n \times n$ identity matrix, and A_3 is a zero matrix.

The corresponding column vector(from the right hand side of the equation) will be returned as well.

Least Square Solver

This function **l_sqr_solver_v4.m** uses the least square constraints method to solve the system of equation. The input to this function is the reduced matrix B & *b'* after applying the function *reduced_matrix_v4.m*. The matrix B is first broke up into block of matrix of the form $\begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix}$. The column vector *b'* is also broke up into two blocks of matrix $\begin{pmatrix} b_e \\ b_l \end{pmatrix}$. There are two steps to the solving of the least square problem. The equations that are to met in the *least square sense* $A_4 x_l = b_l$ are first solved. The exact solutions are then obtained by solving $x_e = b_e - A_2 x_l$.

This function returns the least square solutions x_l and the exact solutions are x_e.

Generate Statistics

The script **test_stat.m** is first called in *test_v4_original_latext.m*. It is used to generate **sum of squares of errors SSE**, **variance** s^2, the **covariance matrix**, the **prediction interval & confidence estimates**. Three other scripts (*print_stat_output,print_stat_output1,print_estimates*) that reside within this script are used for the printing of output. Lastly, the script also plots a scatter diagram.

VITA

Chee Loong Ng

Work Experiences

- Lecturer, Temasek Engineering School, Temasek Polytechnic, Singapore, 2000 - Present
- Administrative Officer, Temasek Engineering School, Temasek Polytechnic, Singapore, 1997 - 2000
- Academic Affairs Officer, Registrar's Office, Temasek Polytechnic, Singapore, 1994 - 1997

Education

- Master of Science, Computer Science, Texas A&M University, 2001 - 2003
- Bachelor of Science, Mathematics, National University of Singapore, 1991 - 1994

Contact Addresses

- *Work:* Department of Computer Engineering & Info-communications, Temasek Engineering School, Temasek Polytechnic, 21 Tampines Avenue 1, Singapore 529757, Singapore
- *Permanent mailing address:* 439 Ang Mo Kio Ave 10, #15-1321, Singapore 560439, Singapore
- *email:* ngcl@tp.edu.sg

Interests

- Info-communications related technologies
- Issues related to the well-being of mankind