Curve Fitting Software

Malavika Srinivasan

October 10, 2018

1 Revision History

Date	Version	Notes
Sep 28, 2018	1.0	0^{th} draft by Malavika
Oct 2, 2018	1.1	1^{st} draft by Malavika

2 Reference Material

This section records information for ease of reference. The information includes the units, symbols and abbreviations used in this document.

2.1 Table of Units

This library is designed to work for any set of data, irrespective of their units. So, table of units is **not applicable**.

2.2 Table of Symbols

The table that follows summarizes the symbols used in this document along with their units. The symbols are listed in alphabetical order.

symbol	**unit**	**description**
f		Function
P		Orthogonal matrix
Φ		Basis function
Π		Basis function
$\sum_{min}^{max}$		Summation
r		Residual
T		Transpose operation of matrix
v		Basis function

2.3 Abbreviations and Acronyms

symbol	**description**
A	Assumption
DD	Data Definition
GD	General Definition
GS	Goal Statement
IM	Instance Model
LC	Likely Change
PS	Physical System Description
R	Requirement
CA	Commonality Analysis
CFS	Curve Fitting Software
T	Theoretical Model

2.4 Mathematical Notations

2.4.1 Matrix form

Each data point can be converted to a system of linear equations that can be represented in a matrix form. For instance, let the data be (t_i, y_i) for i $= 0, 1, 2$, then the matrix form is given by,

[A plcae holder to add matrix form explanation —Malavika]

Contents

3 Introduction

Scientific Computation (SC) is the collection of tools, techniques, and theories that are required to solve problems in the field of science and engineering using computer-based mathematical models. The source data for scientific computation problems are often large sets of data from experiments conducted in a laboratory setup. This large set of data (such as time and temperature) is usually complex to analyze and require segmenting and curve-fitting. Curve fitting is the process of constructing a curve, or mathematical function, that has the best fit to a series of data points possibly subject to constraints.

The following section provides an overview of the Commonality Analysis (CA) for a family of softwares developed to fit a set of data points. The developed program will be referred to as Curve Fitting Software(CFS). This section explains the purpose of this document, the scope of the family, the organization of the document and the characteristics of intended reader.

3.1 Purpose of Document

The purpose of this document is to analyze the commonalities and variabilities between the family of softwares produced as part of this library. The commonalities are the elements which are true for each member which is a part of this software family. For instance, curve fitting is common for all the softwares of this family. The variabilities are the elements which may vary between each software member of this family. For instance, the technique used to fit the curve such as interpolation, regression, smoothing are examples of variabilities.

3.2 Scope of the Family

The scope of CFS is restricted to finding the best fit parameters of linear systems using methods of interpolation and regression discussed in chapter 3 and 7 of Heath (1997) . To reduce the complexity of CFS, we use only polynomials and piecewise polynomials as our basis functions for interpolation. In regression, we restrict ourselves to over determined linear systems and least squares methods to minimize the residual.

3.3 Characteristics of Intended Reader

The reader of this document is expected to know the basic concepts of curve fitting. However to obtain a deeper understanding of this document, the reader is expected to be proficient in elementary linear algebra.

3.4 Organization of Document

The organization of this document follows the template for a Commonality analysis (CA) for scientific computing software proposed by Dr.Spencer Smith. The sections of this document are grouped into commonalities and variabilities of the CFS. The presentation starts with the analyzing commonalities between the softwares of CFS which will contain goals, theories, and definitions which will be common for the family members of CFS. Then the variabilities of the family members is presented which will contain the assumptions, instance models and the output. The goal statements are refined to the theoretical models, and theoretical models to the instance models.

4 General System Description

This section identifies the interfaces between the system and its environment, describes the potential user characteristics and lists the potential system constraints.

4.1 Potential System Contexts

Figure 1 shows the system context. A circle represents an external entity outside the software, the user in this case. A rectangle represents the software system itself (CFS). Arrows are used to show the data flow between the system and its environment.

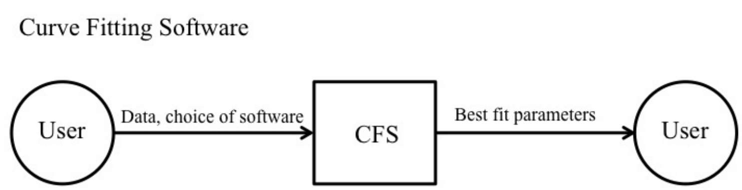

Figure 1: System Context

- User Responsibilities:

- Enter input data points to be fit.
- Input whether to use interpolation or regression.
- Input interpolating method or regression method based on their choice in the previous step.

- CFS Responsibilities:
  - Detect data type mismatch, such as a string of characters instead of a number.
  - Compute best fit parameters.

4.2 Potential User Characteristics

The end user of CFS should have an understanding of undergraduate Level Regression and Interpolation.

4.3 Potential System Constraints

Not Applicable.

5 Commonalities

5.1 Background Overview

Curve fitting is the process of constructing a curve, or mathematical function, that has the best fit to a series of data points possibly subject to constraints. The figure below(2) shows the best fit curve through a series of points.

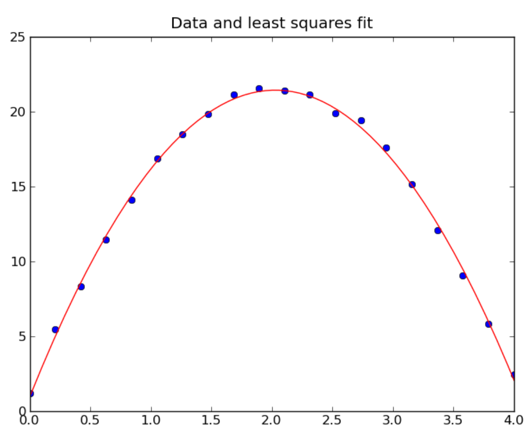

Figure 2: Typical example of a curve fitting process

A fit for the data can be obtained by different methods like interpolation, data smoothing and regression.

5.1.1 Interpolation

Interpolation is a process of fitting a function to given data so that the function has some values as the given data.

5.1.2 Smoothing

Smoothing is a process of creating an approximating function that attempts to capture important patterns in the data, while leaving out noise or other rapid phenomena.

5.1.3 Regression

Regression analysis is the process of finding the best fit parameters for a regression model for a given set of data points and thus obtain a curve through a set of data points.

5.2 Terminology and Definitions

This subsection provides a list of terms that are used in the subsequent sections and their meaning, with the purpose of reducing ambiguity and making it easier to correctly understand the requirements:

- Basis functions: A basis function is an element of a particular basis for a function space. Every continuous function can be represented as a linear combination of basis functions.
- Transpose: This is a mathematical operation on a matrix by which the rows of a matrix becomes its columns and vice versa.
- Cholesky Factorization: This is a particular form of this factorization which is commonly used to solve the normal equations that characterize the least squares solution to the overdetermined linear system.
- Orthogonality: Orthogonality is the generalization of the notion of perpendicularity.
- Interpolant: An interpolant is a function $L = L(t)$ which agrees with a particular function f at a set of known points $t_0, t_1, t_2, ..., t_n$ and which is used to compute values for f(t) at points $t \neq t_i$ where i $= 0, 1, 2, ..., n$.

5.3 General Definitions

This section collects the equations that will be used in deriving the data definitions, which in turn will be used to build the instance models.

Number	GD1
Label	**General form of linear curve fitting**
SI Units	Not applicable
Equation	$f(t) = \sum_{j=1}^{n} x_j \Phi_j(t)$ where i $= 1, 2, 3...m$
Description	The above equation gives the general form of linear curve fitting where, • (t_i, y_i) is the given set of data points, where i $= 1, 2, 3...m$. • The functions $\Phi(t), \Phi_1(t), ...\Phi_n(t)$ are basis functions whose linear combination yields the curve fitting function f. • x_j is the parameters of the best fit.
Source	Heath (1997)
Ref. By	T3, T4, IM1,IM2, IM3, IM4, IM5, IM6 and IM7

Number	GD2
SI Units	Not applicable
Label	**Overdetermined Linear Systems**
Equation	For a given set of n data points, (t_i, y_i) i $= 1, 2, ...n$ represented in matrix form $Ax = y$ as in T1, where A is an m X n matrix with $m > n$ which means there are more equations than unknowns. A solution to the system x can be computed such that the residual vector, $$r = y - Ax$$ is small. where $i = 1, 2, 3....n$.
Description	The symbols used in this equation are . (t_i, y_i) is the given input data. To find a solution to this problem we use least squares method.(T2)
Sources	Some related information is available at: https://s-mat-pcs.oulu.fi/~mpa/matreng/ematr5_5.htm
Ref. By	T1, IM8, IM9, IM10

5.4 Data Definitions

This section collects and defines all the data needed to build the instance models. The dimension of each quantity is also given.

Number	DD1
Label	**Data definitions of monomial interpolation method.**
Symbol	Φ, A and y.
SI Units	Not applicable
Equation	For the given set of data (t_i, y_i) where j $= 1, 2, ...n$, the below equation gives the basis functions. $$\Phi_j(t) = t^{j-1}$$ The matrix A and y is given by, $$A = \begin{bmatrix} 1 & t_1 & t_1^2 & \dots & t_1^{n-1} \\ 1 & t_2 & t_2^2 & \dots & t_2^{n-1} \\ \vdots & \vdots & \vdots & \vdots & \\ 1 & t_n & t_n^2 & \dots & t_n^{n-1} \end{bmatrix}$$ $$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$
Description	To provide the basis functions and matrices necessary for monomial interpolation IM3. The symbols used in the above equation are $\Phi_j(t)$ which represents the basis function at any t in the data (t_i, y_i). A and y are matrices representing the system of equations through the given set of data points.
Sources	Heath (1997)
Ref. By	IM3

Number	DD2
Label	**Data definitions of Lagrange's interpolation method.**
Symbol	l, A and y.
SI Units	Not applicable
Equation	For the given set of data (t_i, y_i) where j $= 1, 2, ...n$, the below equation gives the basis functions. $$l_j(t) = \frac{\Pi_{k=1,k\neq j}^{n}(t-t_k)}{\Pi_{k=1,k\neq j}^{n}(t_j-t_k)} = \frac{(t-t_1)}{(t_j-t_1)}...\frac{(t-t_{j-1})}{(t_j-t_{j-1})}\frac{(t-t_{j+1})}{(t_j-t_{j+1})}...\frac{(t-t_n)}{(t_j-t_n)}$$ The matrix A(Identity matrix) and y is given by, $$A = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$ $$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$
Description	To provide the basis functions and matrices necessary for lagrange's interpolation IM4. The symbols used in the above equation are $l_j(t)$ which represents the basis function at any t in the data (t_i, y_i). A and y are matrices representing the system of equations through the given set of data points.
Sources	Heath (1997)
Ref. By	IM4

Number	DD3
Label	**Data definitions of Newton interpolation method.**
Symbol	π, A and y.
SI Units	Not applicable
Equation	For the given set of data (t_i, y_i) where j $= 1, 2, ...n$, the below equation gives the basis functions. $$\pi(t) = \Pi_{k=1}^{j-1}(t - t_k)$$ When the limits make it vacuous, the product is taken to be 1. Also, $\pi_j(t) = 0$ for $i < j$. The matrix A and y is given by, A is a lower triangular as $\pi_j(t) = 0$ for $i < j$ with its entries defined by $a_{ij} = \pi_j(t_i)$ $$A = \begin{bmatrix} \pi_0(t_0) & 0 & 0 & \dots & 0 \\ \pi_0(t_1) & \pi_1(t_1) & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ \pi_0(t_n) & \pi_1(t_n) & \pi_2(t_n) & \dots & \pi_n(t_n) \end{bmatrix}$$ $$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$
Description	To provide the basis functions and matrices necessary for Newton's interpolation IM5. The symbols used in the above equation are $\pi_j(t)$ which represents the basis function at any t in the data (t_i, y_i). A and y are matrices representing the system of equations through the given set of data points.
Sources	Heath (1997)
Ref. By	IM5

Number	DD4
Label	**Data definitions of Hermite Cubic interpolation method.**
Symbol	A and y.
SI Units	Not applicable
Equation	The matrix A and y is given by, $$A = \begin{bmatrix} 1 & t_0 & t_0^2 & t_0^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & t_n & t_n^2 & t_n^3 \\ 0 & 1 & 2t_0 & 3t_0^2 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & 2t_n & 3t_n^2 \\ 0 & 0 & 2 & 6t_0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 2 & 6t_n \\ 0 & 0 & 0 & 6 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 6 \end{bmatrix} \text{ and } y = \begin{bmatrix} y_0 \\ \vdots \\ y_0^{(1)} \\ y_1^{(1)} \\ \vdots \\ y_n^{(1)} \\ y_1^{(2)} \\ \vdots \\ y_n^{(2)} \\ y_1^{(3)} \\ \vdots \\ y_n^{(3)} \end{bmatrix}$$
Description	To provide the matrices necessary for Hermite cubic interpolation IM6. A and y are matrices representing the system of equations through the given set of data points.
Sources	Some related information can be obtained at: https://coast.nd.edu/jjwteach/www/www/30125/pdfnotes/lecture5_9v14.pdf
Ref. By	IM6

Number	DD5
Label	**Data definitions of B-Spline interpolation.**
Symbol	$v_i^k(t)$
SI Units	Not applicable
Equation	The basis function for B-Spline is given by: $$v_i^k = \frac{t - t_i}{t_{i+k} - t_i}$$
Description	To provide the basis function for BSpline interpolation IM7.
Sources	Heath (1997)
Ref. By	IM7

Number	DD6
Label	**Matrix Transpose**
Symbol	A^T
SI Units	Not applicable
Equation	$$[A^T]_{ij} = [A]_{ji}$$
Description	The symbols used in the equations are, i, j represents the rows and columns of a matrix. T represents the transpose operation of a matrix.
Sources	Heath (1997)
Ref. By	IM8, IM9

Number	DD7
Label	**Data definitions of Regression using Normal Equations.**
Symbol	$A^T A$
SI Units	Not applicable
Equation	$$A^T A = LL^T$$ $$[A^T]_{ij} = [A]_{ji}$$
Description	The symbols used in the equations are, L represents the lower triangular matrix for Cholesky Factorization. i, j represents the rows and columns of a matrix. T represents the transpose operation of a matrix.
Sources	Heath (1997)
Ref. By	IM8

Number	DD8
Label	**Orthogonality Matrix**
Symbol	P
SI Units	Not applicable
Equation	$$P = (\frac{AA^T}{A^T A})y$$
Description	The symbols used in the equations are, P represents the orthogonal matrix of A. T represents the transpose operation of a matrix.
Sources	Some related information is given at: http://uspas.fnal.gov/materials/05UCB/4_LSQ.pdf
Ref. By	IM10

5.5 Goal Statements

Given the set of data points, the choice of software from CFS and the variabilities of the software the CFS should:

GS1: compute the parameters of the curve which is the best possible fit through the set of data points.

5.6 Theoretical Models

This section focuses on the general equations that CFS is based on.

Number	T1
Label	**Matrix-Vector notation for a system of linear equations.**
Equation	$$Ax = y$$
Description	The above equation gives the general form of system of linear equations in matrix vector notation where, A is a m X n matrix y is a m-vector x is an n-vector
Source	Heath (1997)
Ref. By	IM3, IM4, IM5, IM6, IM8, IM9 and IM10

Number	T2
Label	**Least Squares method.**
Equation	For a given overdetermined system of equations represented in matrix form as in T1, a best fit solution x is found such that, $$\min_x \sum_{i=1}^{m} (y_i - f(t_i, x))^2$$
Description	The above equation gives the definition of best fit solution in least squares method. A is a m X n matrix y is a m-vector x is an n-vector
Source	Heath (1997)
Ref. By	IM8, IM9 and IM10

Number	T3
Label	**General form of Interpolation**
Equation	$f(t_i) = \sum_{j=1}^{n} x_j \Phi_j(t) = y_i$ where i $= 1, 2, 3...m$
Description	The above equation gives the general form of interpolating function where, • (t_i, y_i) is the given set of data points, where i $= 1, 2, 3...m$. • The functions $\Phi(t), \Phi_1(t), ...\Phi_n(t)$ are basis functions whose linear combination yields the interpolating function f whose value at a given t_i is y_i. • x_j is the parameters of the best fit. • y_i is the given data point at t_i.
Source	Heath (1997)
Ref. By	GD1, IM1, IM2, IM3, IM4, IM5, IM6, IM8, IM9 and IM10

Number	T4
Label	**General form of Linear Regression using Least squares method and best fit in least squares.**
Equation	For i $= 1, 2, 3...m$, general form of regression is given by $$f(t_i) = \sum_{j=1}^{n} x_j \Phi_j(t)$$ Since we deal with only linear problems(3.2), Φ_j depend only on t. Hence $f(t_i, x)$ can be expressed as, $$f(t_i, x) = x_1 + x_2 t + x_3 t^2 + ... + x_n t^{n-1}.$$ To find x_i using Least squares method as in T2
Description	The above equation gives the general form of linear regression function and the definition of best fit in least squares method. (t_i, y_i) is the given set of data points, where i $= 1, 2, 3...m$. The functions $\Phi(t), \Phi_1(t), ...\Phi_n(t)$ are basis functions whose linear combination yields the function f whose value at any t is the best fit to the given data. x_j is the parameters of the best fit.
Source	Heath (1997)
Ref. By	GD1, IM8, IM9 and IM10

6 Variabilities

6.1 Assumptions

A1: For the given 'n' number of data points (t_i, y_i) where i $= 0, 1, 2, ...n$, n ≥ 2.(R1)

A2: For the given 'n' number of data points (t_i, y_i) where i $= 0, 1, 2, ...n$, $t_i < t_{i+1}$.(IM6,IM7)

A3: For the given set of data points (t_i, y_i) where i $= 0, 1, 2, ...n$, there exists an interpolant.(IM1, IM2, IM3, IM4, IM5, IM6 and IM7).

A4: For the given set of data points (t_i, y_i), there exists a least squares solution.(IM8, IM9 and IM10)

A5: We assume only linear basis functions for B-Splines. (IM7)

6.2 Calculation

6.2.1 Instance Model

This section transforms the goal of CFS defined in the Section 5.5 into one which is expressed in mathematical terms. It uses concrete symbols defined in Section 5.4 to replace the abstract symbols in the models identified in the Sections 5.6 and 5.3.

Number	IM1
Label	**Polynomial Interpolation**
Input	A set of n data points, (t_i, y_i) i $= 1, 2, ...n$
Output	$P_{n-1}(t)$ is a polynomial of degree atmost $n-1$ such that $$P_{n-1}(t_i) = y_i$$ where $i = 1, 2, 3....n$
Description	The symbols used in this equation are P_{n-1} which represents the polynomial of degree at the most $n-1$. (t_i, y_i) is the given input data where no two t_i are the same. Polynomial interpolation has different variabilities which are discussed as separate instance models.
Sources	Some related information is available at Wikipedia
Ref. By	T3, IM3, IM4, IM5, and R5

Number	IM2
Label	**Piecewise Interpolation**
Input	A set of n data points, (t_i, y_i) i $= 1, 2, ...n$ where $x_{i+1} > X_i$.
Output	$P(t) = L_i(t)$ is a piecewise linear function where L_i is gievn by the following equation. $$L_i(t) = y_i + \frac{y_{i+1} - y_i}{t_{i+1} - t_i}(t - t_i)$$ where $i = 1, 2, 3....n$. The degree of the polynomial can be varied as per requirement in each interval.
Description	The symbols used in this equation are $P(t)$ which represents the piecewise polynomial of degree 1. (t_i, y_i) is the given input data where no two t_i are the same. Piecewise interpolation has different variabilities which are discussed as separate instance models.
Sources	Some related information is available at: http://www.math.umd.edu/~petersd/460/spline.pdf
Ref. By	T3, IM6, IM7 and R5

Number	IM3
Label	**Monomial Interpolation**
Input	(t_i, y_i) i $= 1, 2, ...n$ see DD1, from which the basis function of the monomial interpolation is obtained. see DD1, from which the definition of matrix A and b are obtained.
Output	Find $P_{n-1}(t)$ such that for the given input (t_i, y_i), $$P_{n-1}(t_i) = x_1 + x_2 t_i + x_3 t_i^2 + x_4 t_i^3 + ...x_n t_i^{n-1} = y_i$$ Where $x_1, x_2, x_3 ... x_n$ can be obtained by representing the above system of linear equations in matrix form $Ax = b$ as in T1 and solving for x using the equation below. $$Ax = \begin{bmatrix} 1 & t_1 & t_1^2 & \dots & t_1^{n-1} \\ 1 & t_2 & t_2^2 & \dots & t_2^{n-1} \\ \vdots & \vdots & \vdots & \vdots & \\ 1 & t_n & t_n^2 & \dots & t_n^{n-1} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = y$$
Description	To determine the interpolating polynomial p_{n-1} for a given set of input points. The symbols used in this equation are P_{n-1} which represents the polynomial of degree at the most $n - 1$. $x_1, x_2 ... x_n$ are best fit parameters.
Sources	Some related information is available at: Heath (1997)
Ref. By	T3, GD1, R6

Number	IM4
Label	**Lagrange's Interpolation**
Input	(t_i, y_i) i $= 1, 2, ...n$ see DD2, from which the basis function of the lagrange's interpolation is obtained. see DD2, from which the definition of matrix A and b are obtained.
Output	Find $P_{n-1}(t)$ such that for the given input (t_i, y_i), $$P_{n-1}(t) = y_1 l_1(t) + y_2 l_2(t) + \dots y_n l_n(t).$$ Where $x_1, x_2, x_3 ... x_n$ can be obtained by representing the above system of linear equations in matrix form $Ax = b$ as in T1 and solving for x using the equation below. $$Ax = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = y$$
Description	To determine the interpolating polynomial p_{n-1} for a given set of input points. The symbols used in this equation are P_{n-1} which represents the polynomial of degree at the most $n - 1$. $x_1, x_2 ... x_n$ are best fit parameters.
Sources	Some related information is available at: Heath (1997)
Ref. By	T3, GD1, R6

Number	IM5
Label	**Newton's Interpolation**
Input	(t_i, y_i) i $= 1, 2, ...n$ see DD3, from which the basis function of the Newton's interpolation is obtained. see DD3, from which the definition of matrix A and b are obtained.
Output	Find $P_{n-1}(t)$ such that for the given input (t_i, y_i), $P_{n-1}(t) = x_1 + x_2(t - t_1) + x_3(t - t_1)(t - t_2) + ...x_n(t - t_1)(t - t_2)...(t - t_{n-1})$ Where $x_1, x_2, x_3...x_n$ can be obtained by representing the above system of linear equations in matrix form $Ax = y$ as in T1 and solving for x using the equation below. $$Ax = \begin{bmatrix} \pi_0(t_0) & 0 & 0 & \dots & 0 \\ \pi_0(t_1) & \pi_1(t_1) & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ \pi_0(t_n) & \pi_1(t_n) & \pi_2(t_n) & \dots & \pi_n(t_n) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = y$$
Description	To determine the interpolating polynomial p_{n-1} for a given set of input points. The symbols used in this equation are P_{n-1} which represents the polynomial of degree at the most $n - 1$. $x_1, x_2...x_n$ are best fit parameters.
Sources	Some related information is available at: Heath (1997)
Ref. By	T3, GD1, R6

Number	IM6
Label	**Hermite Cubic Interpolation**
Input	(t_i, y_i) where i $= 1, 2, ...n$ m derivatives for the n data points above (where $m = 3$ as a cubic polynomial can have only 3 derivatives) represented by $(t_i^{(m)}, y_i^{(m)})$ where i $= 1, 2, ...n$. see DD4, from which the definition of matrix A and b are obtained.
Output	Find $P(t)$ such that for the given input (t_i, y_i), $$P(t) = x_0 + x_1 t + x_2 t^2 + x_3 t^3$$ Where x_0, x_1, x_2, x_3 can be obtained by representing the above system of linear equations in matrix form $Ax = y$ as in T1 and solving for x using the equation below. $$Ax = \begin{bmatrix} 1 & t_0 & t_0^2 & t_0^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & t_n & t_n^2 & t_n^3 \\ 0 & 1 & 2t_0 & 3t_0^2 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & 2t_n & 3t_n^2 \\ 0 & 0 & 2 & 6t_0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 2 & 6t_n \\ 0 & 0 & 0 & 6 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ x_{n(m+1)} \end{bmatrix} = \begin{bmatrix} y_0 \\ \vdots \\ y_0^{(1)} \\ y_1^{(1)} \\ \vdots \\ y_n^{(1)} \\ y_1^{(2)} \\ \vdots \\ y_n^{(2)} \\ y_1^{(3)} \\ \vdots \\ y_n^{(3)} \end{bmatrix} = y$$
Description	To determine the interpolating polynomial p for a given set of input points. The symbols used in this equation are P which represents the polynomial of degree at the most $n(m+1) - 1$. $x_1, x_2 ... x_{n(m+1)}$ are best fit parameters.
Sources	Some related information can be obtained at: https://coast.nd.edu/jjwteach/www/www/30125/pdfnotes/lecture5_9v14.pdf
Ref. By	T3, GD1, R7

Number	IM7
Label	**B-Spline Interpolation**
Input	(t_i, y_i) i $= 1, 2, ...n$ such that $t_0 < t_1 < ...t_n$ see DD5, from which the basis function for B-Splines is obtained.
Output	Find $B_i^k(t)$ such that for the given input (t_i, y_i), $$B_i^k(t) = v_i^k(t)B_i^{k-1}(t) + (1 - v_{i+1}^k(t))B_{i+1}^{k-1}(t)$$ for $k > 0$.
Description	To determine B-Spline B_i^k for a given set of input points. The symbols used in this equation are B_i^k which represents a piecewise polynomial of degree k. Important properties of B-Splines includes: • For $t < T_i ort > t_{i+k+1}, B_i^k(t) = 0$ • For $t_i < t < t_{i+k+1}, B_i^k(t) > 0$ • For all t, $\sum_{i=-\infty}^{\infty} B_i^k(t) = 1$ • For $k \geq 1 B_i^k$ is $k - 1$ times continously differentiable.
Sources	Some related information is available at: Heath (1997)
Ref. By	T3, GD1, R7

Number	IM8
Label	**Normal Equations**
Input	A set of data points (t_i, y_i) for i $= 1, 2, ...n$ represented as a overdetrmined system of equations as in T1. The matrices $A^T A$ and A^T from DD7 and DD6
Output	x_j for j $= 1, 2..m$ by solving the equations below, $$A^T Ax = A^T y$$
Description	The symbols used in the above equations are: T which represents the transpose function of a given matrix.
Sources	Heath (1997)
Ref. By	GD1, GD2, T2, T4 and R8

Number	IM9
Label	**Augmented System**
Input	A set of data points (t_i, y_i) for i $= 1, 2, ...n$ represented as a overdetrmined system of equations as in T1. The matrix A^T from DD6
Output	x_j for j $= 1, 2..m$ by solving the equations below, $$r + Ax = y$$ $$A^T r = 0$$
Description	The equations above give the system to be solved for finding x. The symbols used in the equation are, r - The residual vector.
Sources	Heath (1997)
Ref. By	GD1, GD2, T2, T4 and R8

Number	IM10
Label	**Orthogonal Transformation**
Input	A set of data points (t_i, y_i) for i $= 1, 2, ...n$ represented as a overdetrmined system of equations as in T1. The matrix P from DD8
Output	x_j for j $= 1, 2..m$ by solving the equations below, $$Ax = Py$$
Description	The equation above give the system to be solved for finding x. The symbols used in the equation are, P - The orthogonality matrix.
Sources	Heath (1997)
Ref. By	GD1, GD2, T2, T4 and R8

6.3 Output

7 Requirements

This section provides the functional requirements, the business tasks that the software is expected to complete, and the nonfunctional requirements, the qualities that the software is expected to exhibit.

7.1 Functional Requirements

R1: Read the input the data (t_i, y_i).

R2: Verify the input data to see if only numbers are entered.

R3: Verify the input data to see if A1 and A2 are satisfied.

R4: Read the user's choice of curve fitting method between Interpolation(T3) or Regression(T4).

R5: If interpolation, is chosen in R4, read user's choice of basis function between Polynomial(IM1) or Piecewise(IM2). (A3)

R6: If Polynomial is chosen in R5, read user's choice of basis functions from Monomial(IM3), Lagrange(IM4) or Newton(IM5).(A3)

R7: If Piecewise Polynomial is chosen in R5, read user's choice of interpolating method from Hermite cubic(IM6) or B-Splines(IM7).(A3)

R8: If regression is chosen in R4, read user's choice of regression methods between Normal equations(IM8), Augmented system(IM9) and Orthogonal transformations(IM10).(A4)

R9: Compute best fit parameters.

R10: Display the plot with data(R1) and best fit result from R9.

7.2 Nonfunctional Requirements

7.2.1 Portability Requirements

- CFS shall be able to run on Windows 7/10, Linux, and Mac OSX operating systems.

7.2.2 Verifiability Requirements

- CFS should be verifiable, which means the results should be comparable to Matlab.

7.2.3 Maintainability and Support Requirements

- CFS should be able to accommodate new fitting methods for linear problems.

8 Likely Changes

LC1: The user may want to use smoothing to fit the data.(A??).

9 Traceability Matrices and Graphs

The traceability matrix between data definitions and general definitions is represented below.

	DD1	DD2	DD3	DD4	DD5	DD6	DD7	DD8
GD1	✓	✓	✓	✓	✓	✓	✓	✓
GD2						✓	✓	✓

The traceability matrix between data definitions and Theoretical models is represented below.

	DD1	DD2	DD3	DD4	DD5	DD6	DD7	DD8
T1	✓	✓	✓	✓		✓	✓	✓
T2						✓	✓	✓
T3	✓	✓	✓	✓	✓			
T4						✓	✓	✓

The traceability matrix between Data definitions and Instance models is represented below.

	IM1	IM2	IM3	IM4	IM5	IM6	IM7	IM8	IM9	IM10
DD1			✓							
DD2				✓						
DD3					✓					
DD4						✓				
DD5							✓			
DD6								✓	✓	
DD7									✓	
DD8										✓

The traceability matrix between Theoretical models and Instance models is represented below.

	IM1	IM2	IM3	IM4	IM5	IM6	IM7	IM8	IM9	IM10
T1			✓	✓	✓	✓		✓	✓	✓
T2								✓	✓	✓
T3	✓	✓	✓	✓	✓	✓	✓			
T4								✓	✓	✓

References

Michael T. Heath. *Scientific computing.* Mcgraw-Hill, 1997. ISBN 0071153365 9780071153362.

Wikipedia. Polynomial interpolation. URL https://en.wikipedia.org/wiki/Polynomial_interpolation.

10 Appendix

10.1 Symbolic Parameters