

Data Approximation

Anouk Allenspach, Rodolphe Farrando

December 15, 2017

1 Goal

The goal of this project is to find a polynomial approximation of some data. Several approximation can be made. The goal of the user can be either to find a trend for the or to interpolate the data. Several possibilities are available to interpolate, the first one is a simple polynomial interpolation, where the degree is equal to the number of points minus 1. Then, a piecewise interpolation can be made, this interpolation can be spline or not.

2 Requirements

A gcc compiler (version 4.7.2 or above) **a verifier???** and the CMake software (version 2.6 or above) to link the files are necessary to compile the program. An external library is used to solve linear systems, that is the Eigen library.

3 Method

To approximate the data, the main method that have been used is the least squares method. The method aim to minimize the error between the true value and the estimated one, that is:

$$\min \sum_i (y_i - p(x_i))^2 \quad (1)$$

$p(x_i) = a_0 + a_1 \cdot x_i + \dots + a_m \cdot x_i^m$ being the interpolant polynomial at x_i of degree m . By deriving this equation with respect with all the coefficient of the polynomial, the following linear system is obtained:

$$\begin{bmatrix} \sum_i x_i^0 & \sum_i x_i^1 & \cdots & \sum_i x_i^m \\ \sum_i x_i^1 & \sum_i x_i^2 & \cdots & \sum_i x_i^{m+1} \\ \vdots & \ddots & \ddots & \vdots \\ \sum_i x_i^1 & \cdots & \sum_i x_i^{2m-1} & \sum_i x_i^{2m} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_m \end{bmatrix} = \begin{bmatrix} \sum_i y_i \\ \sum_i y_i x_i \\ \vdots \\ \sum_i y_i x_i^m \end{bmatrix} \quad (2)$$

The linear system $Xa = b$ can be solved by computing $a = X^{-1}b$. For the spline interpolation, <https://www.math.uh.edu/~jingqiu/math4364/spline.pdf> the following paper has been used.

4 Usage of the program

4.1 Input file

The program developed allows the user to give a .csv file with some input inside. Depending on what is in the file, the program will compute what the user has asked. The .csv should have the following form:

data.csv	Type	Approximation Degree
----------	------	----------------------

The *data.csv* file contains the points coordinate. The type is the approximation wanted by the user; it can be either *Data fitting*, *Interpolation*, *Piece-wise Interpolation* or *spline*. Finally, the approximation degree is the degree of the polynomial which approximate the data.

4.2 Predefine Main

The predefined main allows the user to interact with a small program that gives the results wanted. Just compile the program and enter the name file in the console and the program will give the coefficient, the function(s) and the error. To execute the program ***to complete***

4.3 Create its own main

If the user wants to use his own main here are some usage tips. First the Config class will read all the information in the .csv. How to use the different class and methods.

5 Class description

Readfile (class): This class creates an object that instantiate the class Approximation.

Approximation (class): The approximation class has three derived class: Fitting, Interpolation and PieceWiseInterpolation. It will give the coefficients of the polynomial, the display of the function(s) and the error associated.

6 Tests

The series of tests are as follows: Test1: Test2: Test3:

7 Remarks

When the degree of interpolation is high (≥ 7 approximatively), the polynomial is badly condition and the points are not exactly interpolated. Indeed, the condition number of the matrix X in the linear system $Xa = b$ is big. It results difficulties to inverse the matrix and

the solution is biased. The user needs to interpolate few points to have a good result. The same happen for fitting the data, the degree shouldn't be too high otherwise same problems could happen.