Systems Identification - Home Taken Exam

Submitted by Anurag Deepak Shinde (2132290)

January 2023

1 Question 1: Generate your personal dataset, based on your HAN student number, using ExamApp.exe.

With my student number 2132290 I received the dataset DataSet_2132290.xls.

2 Question 2: Read the data from Excel and plot the data in MATLAB (1 point)

The data from the excel file was read in matlab using the xlsread command and the following plot was generated:

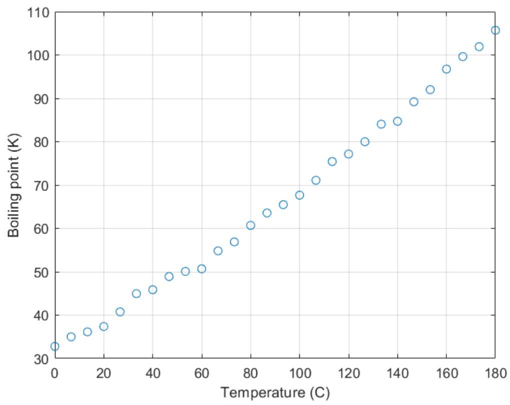

Figure 1: Data plot

3 Question 3: Determine a model (and order) for fitting the data using ordinary linear least squares (OLS) (1 point)

To determine the order of the data, different fits were generated using 'fit' function in Matlab to fit a polynomial to the provided data. Polynomials of the orders 2,3,4,5,6 and exponential model were used to check the fit.

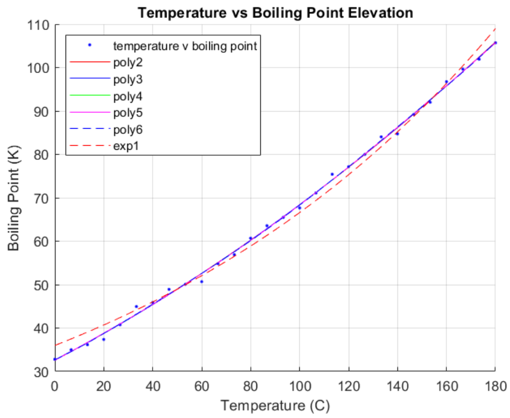

Figure 2: Plot of different fits with data

From the above plot it was difficult to eliminate any fits, therefore the residuals were plotted and it was observed that they are quite similar to each other.

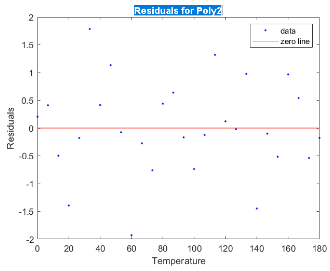

Figure 3: Residuals for Poly2

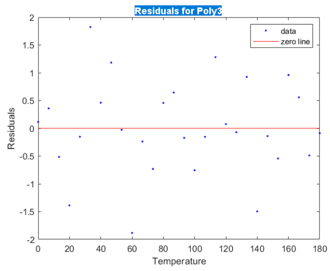

Figure 4: Residuals for Poly3

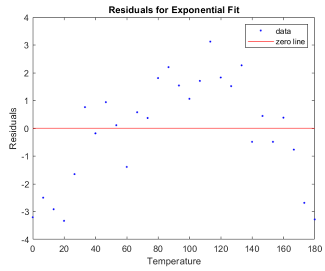

Figure 5: Residuals for Poly4

As the residual plots were not useful therefore examining fits beyond the data range till 300 degrees Celsius.

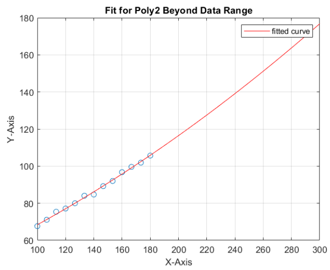

Figure 6: Fit for Poly2 Beyond Data Range

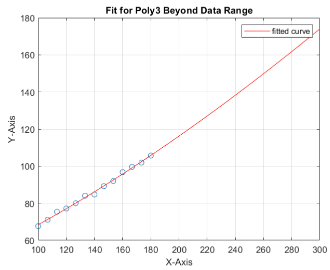

Figure 7: Fit for Poly3 Beyond Data Range

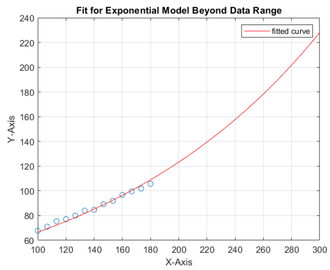

Figure 8: Fit for Exponential Model Beyond Data Range

Even fitting the data beyond the data range was not useful therefore the plots with prediction intervals upto 300 degrees were plotted.

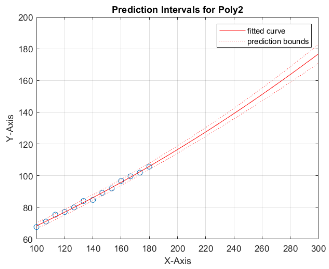

Figure 9: Prediction Intervals for Poly2

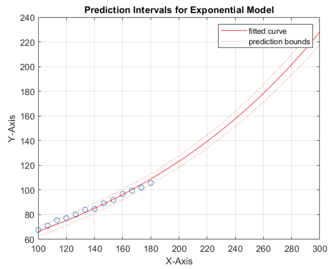

Figure 10: Prediction Intervals for Exponential Model

From the plots of prediction intervals it was observed that the polynomial 2 and the exponential fit might be good fits with the data.

Yet after this the goodness of fit and confidence intervals of all the models were examined and the confidence intervals confirm that the polynomial 2 is the best fit for the provided data as its coefficients do not cross zero while the coefficients of the rest of the models cross zero atleast once.

The 2nd order polynomial fit was then evaluated for new query points and plotted for an extended temperature limits upto 300 degrees Celsius. The result of the evaluation is displayed below which confirms the selection of this fit.

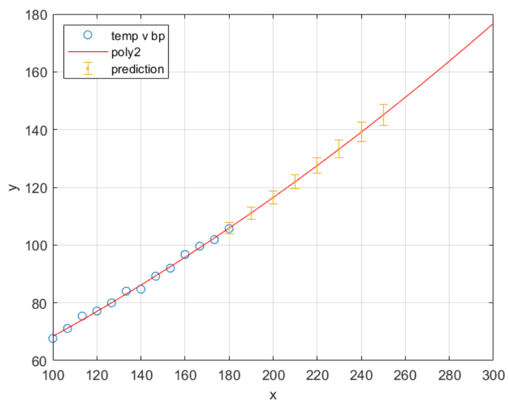

Figure 11: Confidence intervals polynomial 2

4 Using the data and your preferred model construct the design matrix X for the OLS fit

As the chosen model is polynomial of order 2, the design matrix will include column for the second power of the independent variable i.e temperature in this case. Therefore, the design matrix 'X design' will have its first column as all ones (for the intercept), the second column as the temperature, and the third column as the temperature squared, assuming a quadratic model (order 2). The design matrix X can be found in the matlab file.

5 Find the least squares estimator of the parameters P in your model (find $\hat{P}$).

To find the least squares estimator of the parameters $\hat{P}$ in the polynomial model, the design matrix X and the dependent variable vector Y are used. The least squares estimator $\hat{P}$ is calculated by solving the normal equation:

$$\hat{P} = (X^T X)^{-1} XY$$

Therefore, from matlab we have,

$$\hat{p} = \begin{bmatrix} 32.5799 \\ 0.2970 \\ 0.0006 \end{bmatrix}$$

Below given is the OLS estimation plot:

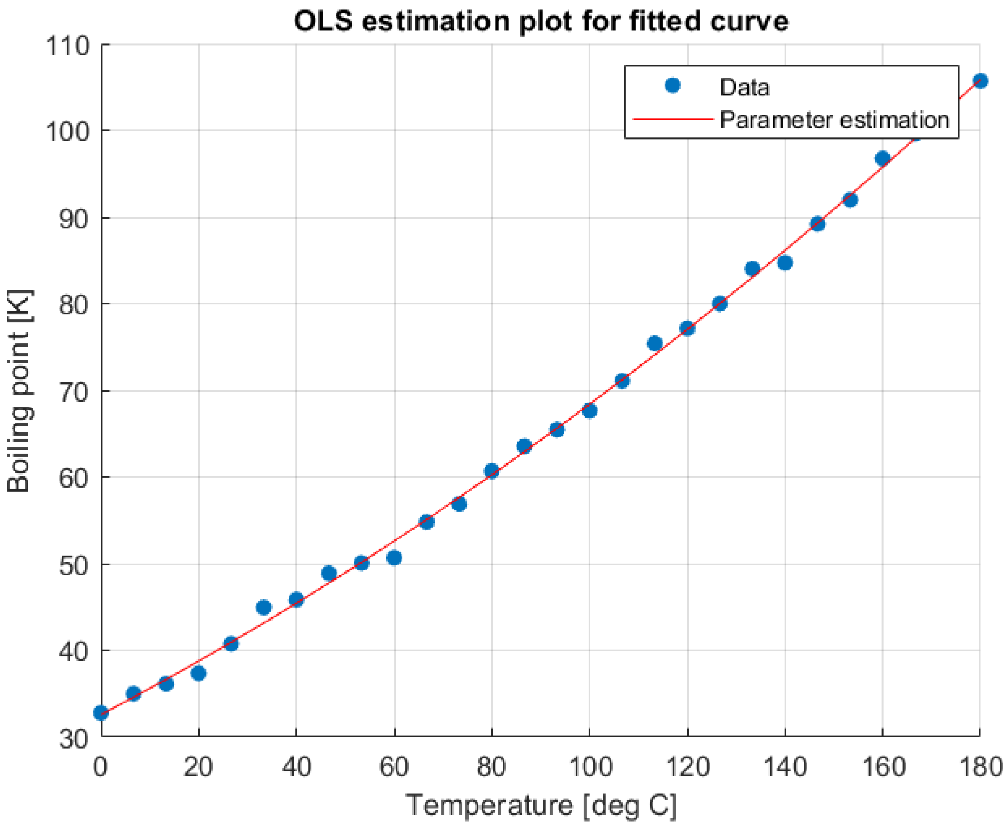

Figure 12: OLS Estimation Plot

6 Determine the variance-covariance matrix of the estimator $\hat{P}$.

The variance-covariance matrix of $\hat{P}$ is given by:

$$Var(\hat{P}) = \sigma^2 (X^T X)^{-1}$$

where σ^2 is the variance of the error term in the model. Since it is usually unknown, it is often estimated using the residual mean square (RMS), which is the sum of squared residuals divided by the degrees of freedom.

Below is the matlab code used for this task:-

```
residuals = Y - X_design * p_hat;
RMS = sum(residuals.^2) / (length(Y) - length(p_hat));
var_cov_matrix = RMS * inv(X_design' * X_design);
```

The variance-covariance matrix is as given below:-

$$\text{var\_cov\_matrix} = \begin{bmatrix} 0.2155 & -0.0047 & 0.0000 \\ -0.0047 & 0.0001 & -0.0000 \\ 0.0000 & -0.0000 & 0.0000 \end{bmatrix}$$

7 Compute the estimate σ_e^2 for the homoskedastic variance σ^2 from the residuals.

The formula to compute σ_e^2 is:

$$\sigma_e^2 = \frac{RSS}{n-m}$$

where, RSS is the sum of the squared residuals n is the number of observations m is the number of estimated parameters, including the intercept.

From matlab we have, $\sigma_e^2 = 0.7707$

8 Calculate the variances and standard deviations in the parameters $\hat{P}$.

To calculate the variances and standard deviations of the estimated parameters $\hat{P}$ in a regression model, the variance-covariance matrix of $\hat{P}$ that was computed earlier was used. The diagonal elements of this matrix represent the variances of the individual parameter estimates, and the standard deviation of each parameter is simply the square root of its variance.

Therefore, from matlab we have,

$$\text{variances} = \begin{bmatrix} 0.2155 \\ 0.0001 \\ 0.0000 \end{bmatrix}$$

$$\text{std\_devs} = \begin{bmatrix} 0.4642 \\ 0.0119 \\ 0.0001 \end{bmatrix}$$

9 Show how the design matrix X can be decomposed by Singular Value Decomposition (SVD)

For a given matrix X, the SVD is a factorization of the form:

$$X = USV^T$$

where, U is an m x m orthogonal matrix,

S is an m x n diagonal matrix with non-negative real numbers on the diagonal, and

V is an n x n orthogonal matrix.

The diagonal entries of S are known as the singular values of X. The columns of U and V are called left-singular vectors and right-singular vectors of X, respectively.

The SVD of the design matrix X was performed in matlab using the 'svd' function as given below:

```
[U, S, V] = svd(X design);
```

10 Solve the linear least squares problem by using the SVD

The SVD of X is: $X = USV^T$ this decomposition is used to solve for $\hat{P}$.

The least squares solution $\hat{P}$ can be found by rearranging the normal equation $X^T X\hat{P} = X^T Y$ and substituting the SVD of X:

$\hat{P} = VS^{-1}U^T Y$

However, since S might be singular or nearly singular its pseudo-inverse is used:

```
sigma_inv = diag(1./diag(S));
p_hat_svd = V * sigma_inv * U' * Y;
```

In matlab in this computation the reciprocal of the non-zero elemnts of S are taken to compute the pseudo-inverse.

This methods yields $\hat{P}$, the estimated parameters of the linear model.

Therefore from matlab we have,

$$\text{p\_hat\_svd} = \begin{bmatrix} 32.5799 \\ 0.2970 \\ 0.0006 \end{bmatrix}$$

Below given is the SVD estimation plot:

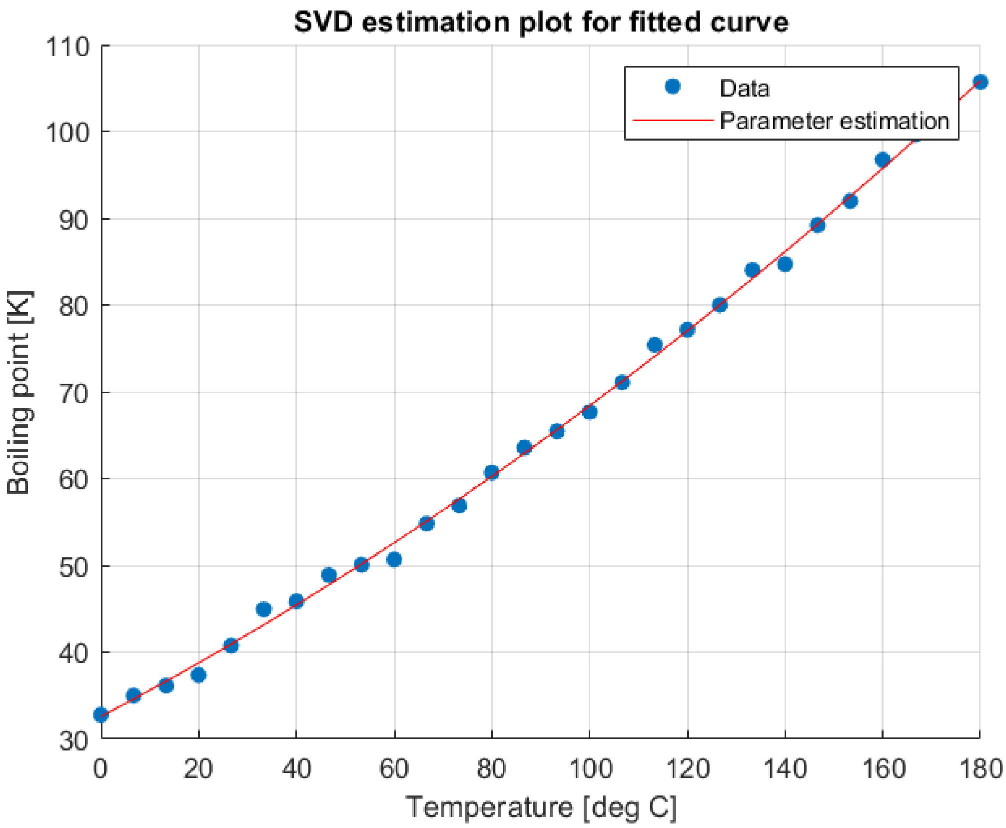

Figure 13: SVD Estimation plot