Perception for Autonomous Robots

Homework 1

Line Fitting using Linear Least Square Techniques

Pranali Desai (116182935) Raghav Nandwani (116321549)

Sanket Goyal (116155144)

Contents

List of Figures

1 Introduction

The pickled files provided are three different data sets on which a line has to be fit using different linear least square techniques. The three basic criteria to be fulfilled in the assignment are:

1. Line fitting using Linear Least Squares
2. Least Square Estimation with Regularization
3. Outliers rejection using RANSAC.

1.1 EigenValues - EigenVectors

After finding the covariance matrix of the datasets, the Eigenvalue and Eigenvectors are found out. Instead of using the direct numpy function to calculate the Eigenvalues and vectors, the values are found out with the help of mathematical functions and equations.

$$S_1 = \begin{bmatrix} 3390.2743 & 1465.6986 \\ 1465.6986 & 1163.6838 \end{bmatrix}, \lambda_1 = [4117.55\ ,\ 436.409]\,, v_{21} = \begin{bmatrix} 0.896 \\ 0.444 \end{bmatrix}, v_{32} = \begin{bmatrix} -0.444 \\ 0.896 \end{bmatrix}$$

$$S_2 = \begin{bmatrix} 3401.1994 & 1034.5587 \\ 1034.5587 & 2166.8373 \end{bmatrix}, \lambda_2 = [3988.686\ ,\ 1579.351]\,, v_{21} = \begin{bmatrix} 0.87 \\ 0.494 \end{bmatrix}, v_{32} = \begin{bmatrix} -0.494 \\ 0.87 \end{bmatrix}$$

$$S_3 = \begin{bmatrix} 3070.7107 & 324.4351 \\ 324.4351 & 2357.5041 \end{bmatrix}, \lambda_3 = [3196.211\ ,\ 2232.004]\,, v_{31} = \begin{bmatrix} 0.933 \\ 0.361 \end{bmatrix}, v_{32} = \begin{bmatrix} -0.361 \\ 0.933 \end{bmatrix}$$

S_1,S_2 and S_3 are the covariance matrices for data1, data2 and data3 respectively

Eigenvalues The co-variance matrix calculated is subtracted from Lambda times identity matrix followed by the determinant of the resulting matrix. This quadratic equation is then solved for the value of Lambda in Python.

Eigenvectors To find the Eigenvectors, the calculated values of Lambda are substituted into the above matrix followed by the representation of one element 'v2' in terms of 'v1'. The next step is to normalize the Eigenvectors to eliminate the constant value added in the above step to maintain the ratio.

1.2 Least Square and Total least square

1.2.1 Least Square

1. Let the data points in the data set be $[(x_1, y_1), (x_2, y_2), x_3, y_3) .... (x_n, y_n)]$

2. Write it in the form $XB = Y$ such that

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ . \\ . \\ y_n \end{bmatrix}, X = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ x_3 & 1 \\ . & . \\ . & . \\ x_n & 1 \end{bmatrix}, B = \begin{bmatrix} m \\ b \end{bmatrix}$$

3. As X is not a square matrix so we have to calculate the pseudo inverse of X to calculate B.

$$B = X^{+}Y \ , \ X^{+} = \left(X^T X\right)^{-1} X^T$$

   where X^{+} is pseudo inverse of X.

4. With the value of m and b we plot the line in form of $y = mx + b$

1.2.2 Total Least Square

Calculate the mean of the data set

$$\bar{x} = \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right) \ , \ \bar{y} = \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)$$

Calculate U and then $U^T U$ such that

$$U = \begin{bmatrix} x_1 - \bar{x} & y_1 - \bar{y} \\ x_2 - \bar{x} & y_2 - \bar{y} \\ . & . \\ . & . \\ x_n - \bar{x} & y_n - \bar{y} \end{bmatrix}, \ U^T U = \begin{bmatrix} \sum_{i=1}^{n} (x_i - \bar{x})^2 & \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) \\ \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) & \sum_{i=1}^{n} (y_i - \bar{y})^2 \end{bmatrix}$$

Now line is plotted in the form such that $ax + by - d = 0$. To calculate a, b and d, eigenvector components of the least eigenvalue of matrix $U^T U$ will give the value of a and b. the value of d is calculated such that

$$d = a\bar{x} + b\bar{y}$$

1.3 Outlier Rejection Technique

Two types of outlier detection techniques have been used:

1. Regularization
2. RANSAC

1.3.1 Regularization

In this method, the least square fitting concept has been used with regularization. After generating the line, we try to find a set of inliner using the threshold. Here we see that the output is vague and there are way too many outliers. For eg: In data set 1 we get only around 99 inliner with regularized outlier detection, but when we implement using RANSAC inlner percent is a good number around 174.

Disadvantages of Regularization -

1. Number of inliner are not satisfactory.
2. No better results even after tuning like RANSAC.

1.3.2 RANSAC

The RANdom SAmple Consensus (RANSAC) algorithm generated by Fischler and Bolles is a generalized parameter estimation method devised to cope with a high proportion of outliers in the input data. This has also been applied to the given data points.

Algorithm for RANSAC

1. Select randomly 2 points to determine the model parameters.
2. Solve for the parameters of the model.
3. Determine how many points are within the threshold.
4. If the fraction of inliers over the total number points in the set passes a predefined threshold, re-estimate the model parameters utilizing all the known inliers and terminate.
5. Otherwise, repeat steps 1 to 4.

Once the outliers have been detected, they have been eliminated and line fitting using Vertical Fit had been applied on it.

Advantages for RANSAC

1. It is a very robust approach to finding out outliers.
2. A very quick approach when the number of data set is large.
3. Very modular approach.
4. Thus it is the most general technique for outlier detection.

Due to its varied advantages and parameter tuning capabilities, RANSAC has been chosen as the outlier rejection technique for all three data sets.

The Outlier Rejection Technique RANSAC has a number of drawbacks resulting in an increase of number of iterations.

1. For dataset 1, the number of inliers is 179 compared to 126 in dataset 2 and only 73 in dataset 3. Repeated iterations with respect to required percentage of inliers have to be performed followed by re-running the program again to achieve the best fit line.
2. The parameters required to tune RANSAC require the knowledge of the threshold and the percentage of inliers in the system. Apart from this, the parameter to construct the line, which is minimum number of points is also required.

Difficulties Faced while implemnting the problem statement

1. To find eigen values without using any inbuilt functions.
2. Finding the equations for TSL and OSL
3. Understanding the concept of fitting a line after RANSAC.
4. Implementing parameters and tuning to find the best set of inliers.

1.4 Results

Fig 1, Fig 2 and Fig 3 shows the eigenvector of the covariance matrix of data1, data2 and data3 respectively.

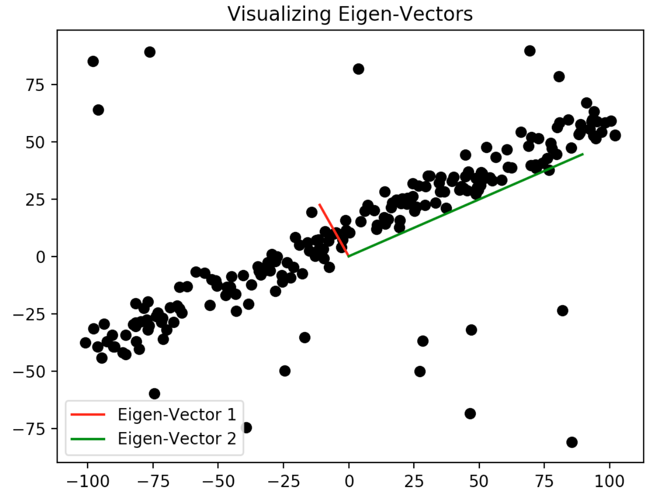

Figure 1: Eigen-Vectors: Dataset 1

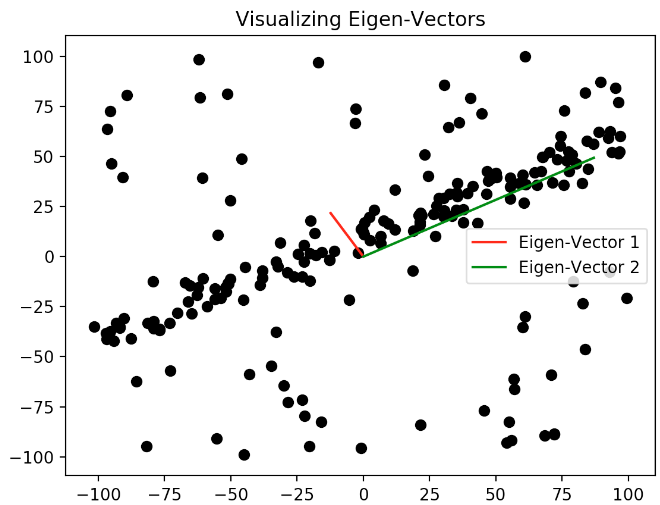

Figure 2: Eigen-Vectors: Dataset 2

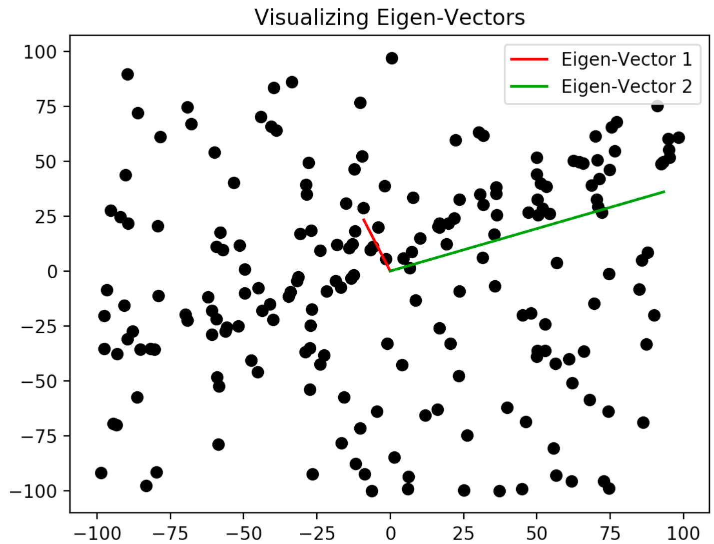

Figure 3: Eigen-Vectors: Dataset 3

Fig 4, Fig 5 and Fig 6 shows the results of Least Square - vertical fit (LS-V), Least Square - regularisation (LS-R) and Total least square (TLS) plot of data1, data2 and data3 respectively.

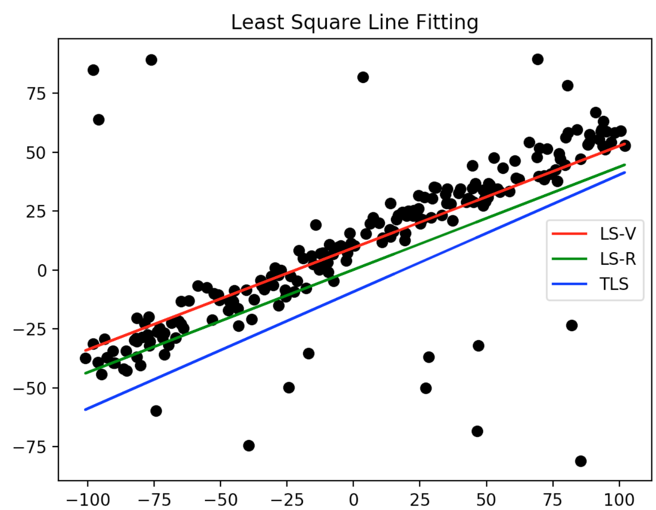

Figure 4: Line Fitting using Least Square: Dataset 1

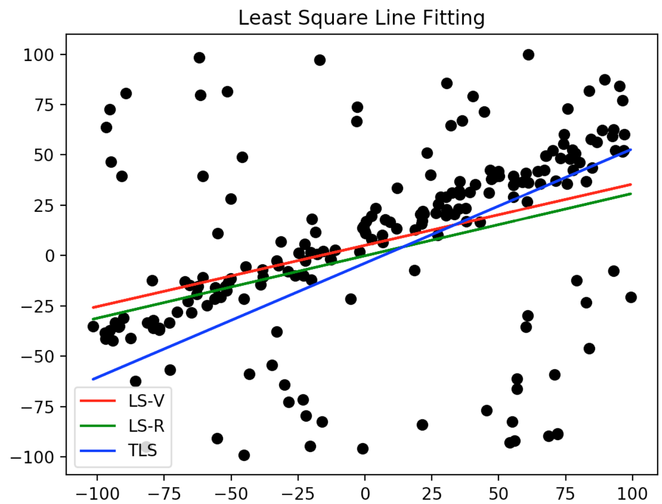

Figure 5: Line Fitting using Least Square: Dataset 2

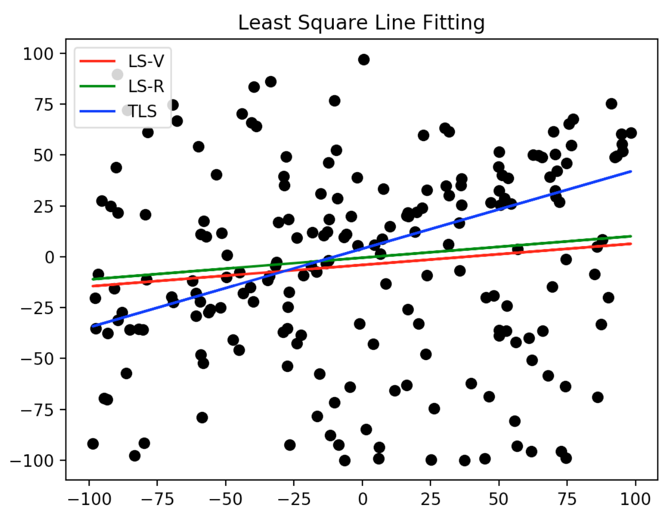

Figure 6: Line Fitting using Least Square: Dataset 3

Fig 7, Fig 8 and Fig 9 are the data plots without any outlier i.e. only inlier for data1, data2 and data3 respectively.

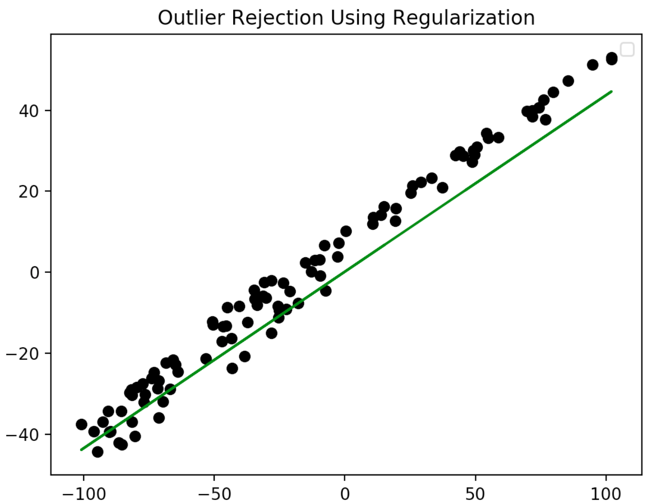

Figure 7: Outlier Rejection using Regularization: Dataset 1

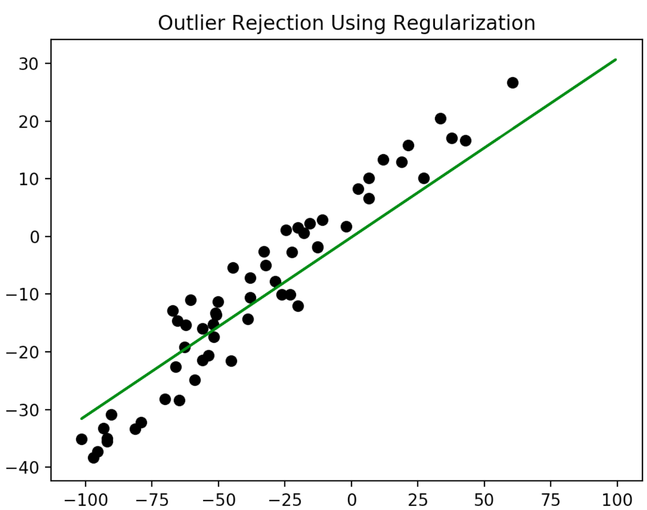

Figure 8: Outlier Rejection using Regularization: Dataset 2

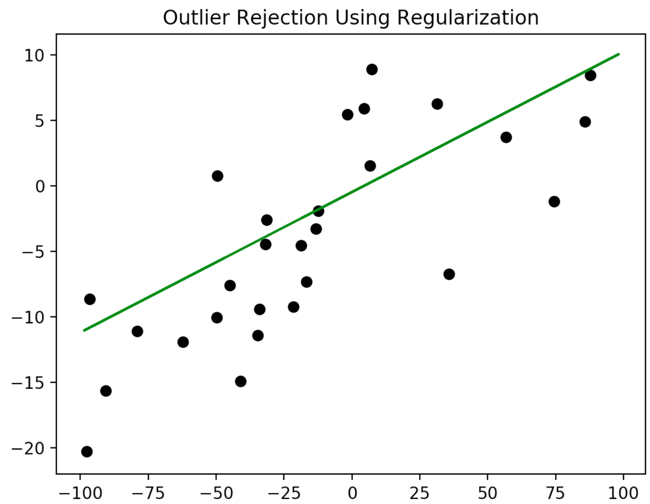

Figure 9: Outlier Rejection using Regularization: Dataset 3

RANSAC plots of data1, data2 and data3 in Fig 10, Fig 11 and Fig 12.

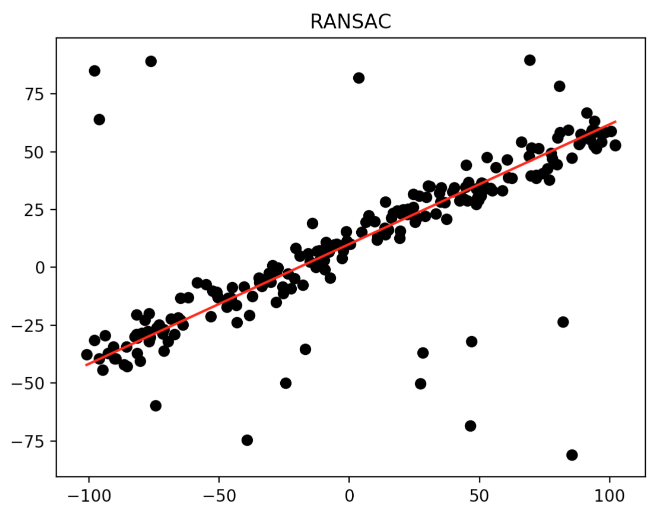

Figure 10: Implementing RANSAC: Dataset 1

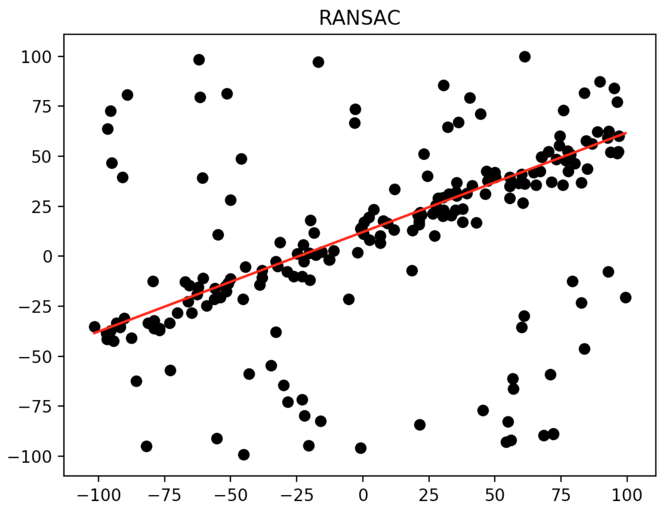

Figure 11: Implementing RANSAC: Dataset 2

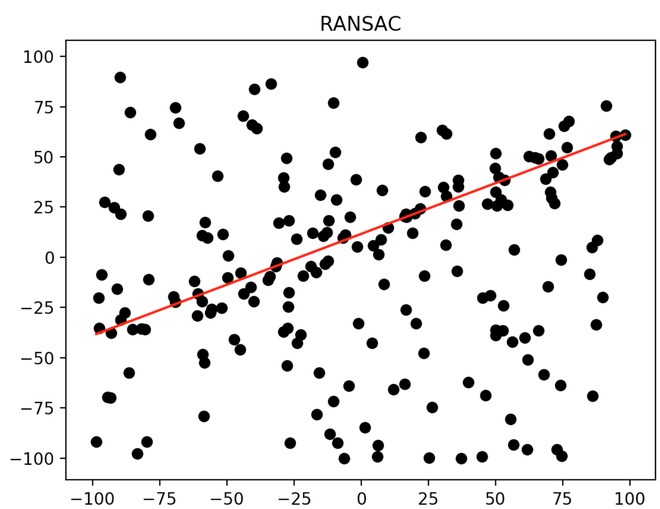

Figure 12: Implementing RANSAC: Dataset 3

RANSAC plots without outliers represented in Fig 13, Fig 14 and Fig 15.

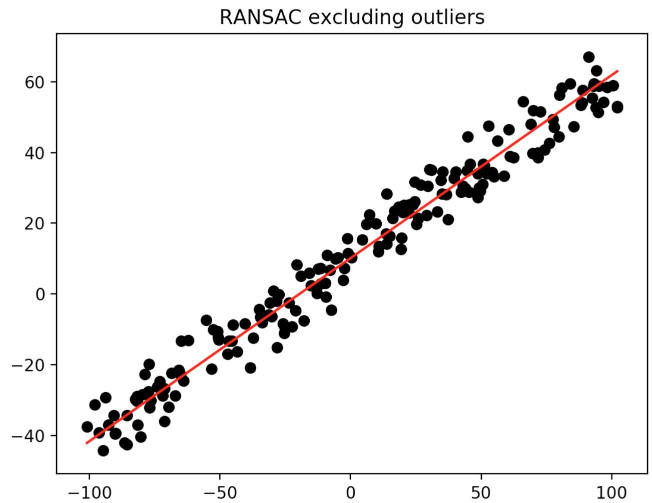

Figure 13: RANSAC excluding Outliers: Dataset 1

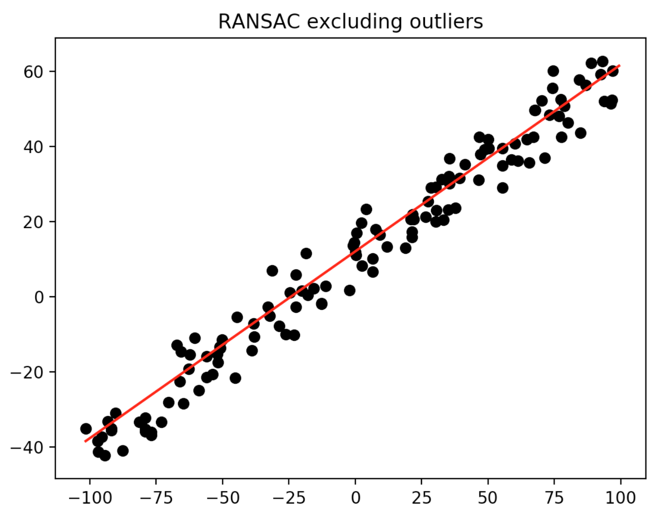

Figure 14: RANSAC excluding Outliers: Dataset 2

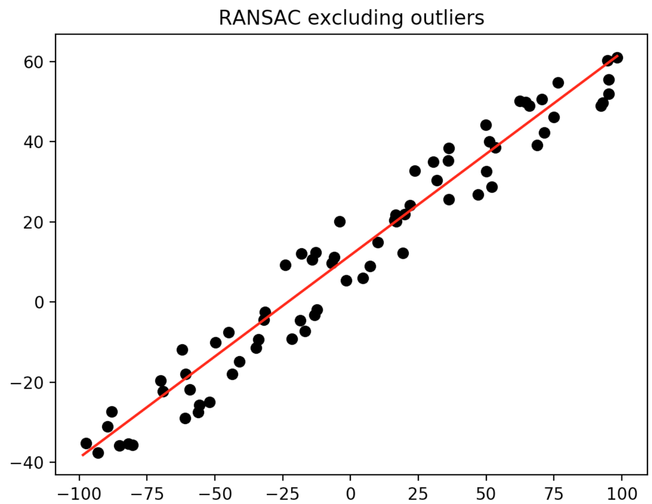

Figure 15: RANSAC excluding Outliers: Dataset 3

Line fitting after the RANSAC operation using only the inlier data given by Fig 16, Fig 17 and Fig 18.

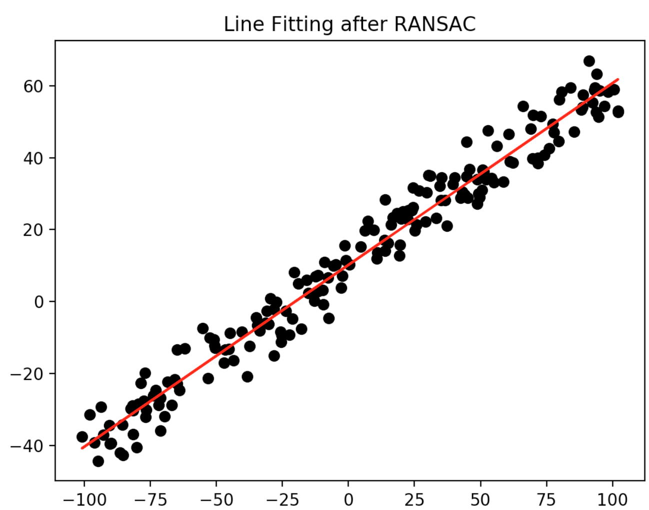

Figure 16: Line Fitting after RANSAC: Dataset 1

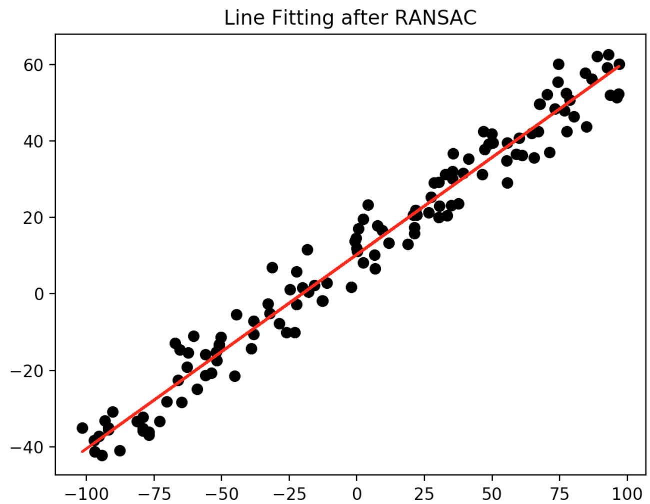

Figure 17: Line Fitting after RANSAC: Dataset 2

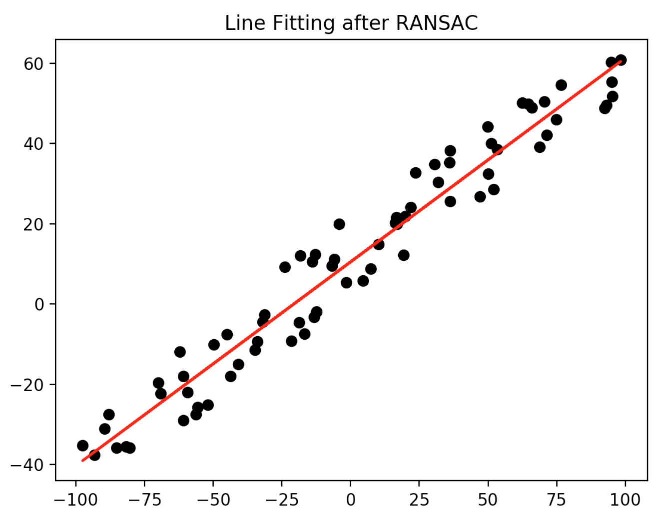

Figure 18: Line Fitting after RANSAC: Dataset 3