ENPM 673 – HW 1

Name: Hrishikesh Tawade
UID: 116078092

Name: Krishna Bhatu
UID: 116304764

Name: Kapil Rawal
UID:116133357

1) Understanding of Eigen Vectors and Eigen Values

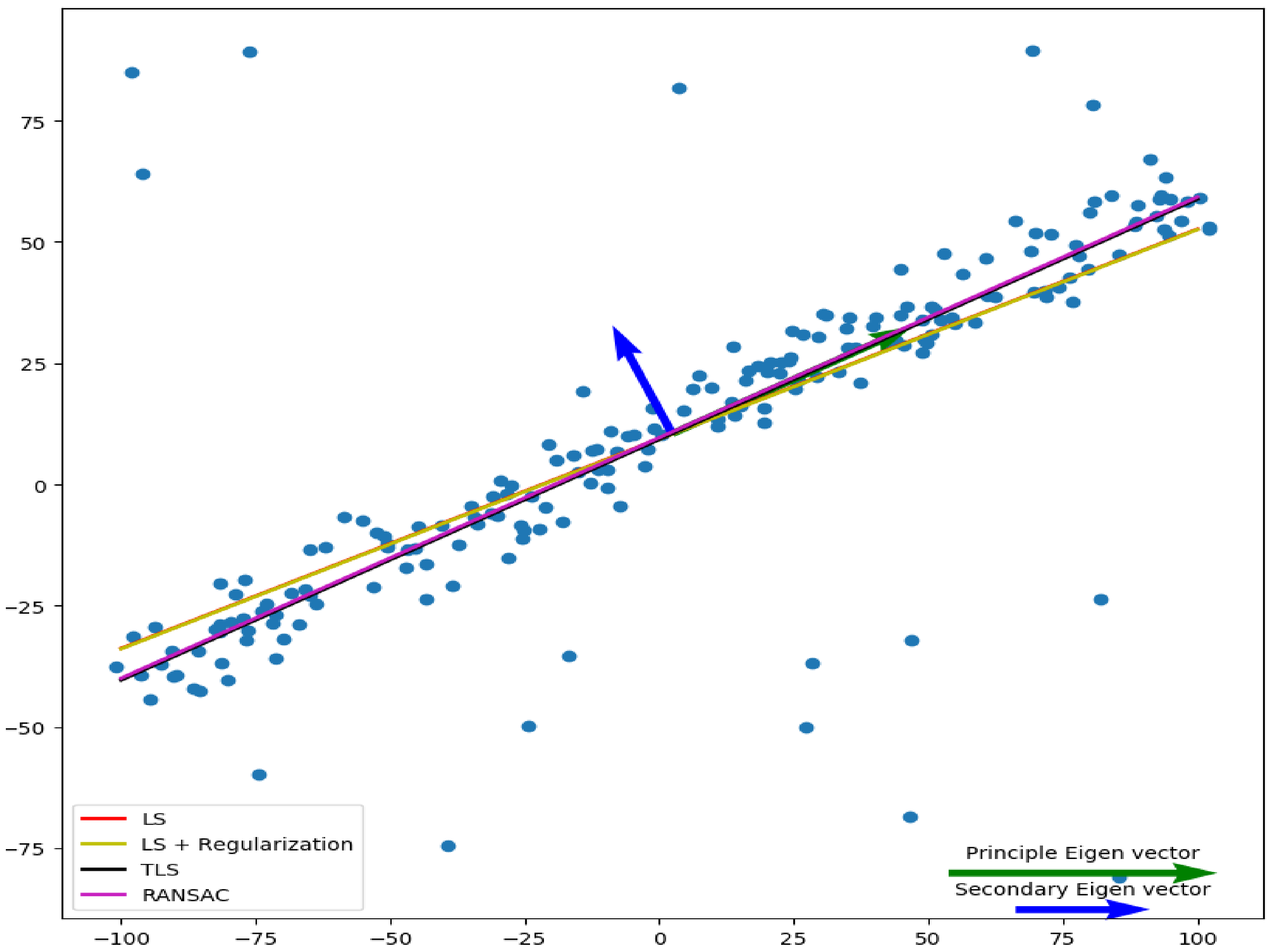
- An eigen vector is a vector which on linear transformation undergoes only scaling and no change in direction and the scalar value is called eigen value.
- In general, the eigenvector $\vec{v}$ of a matrix A is the vector for which the following holds:

$$A\vec{v} = \lambda\vec{v}$$

  Here $\vec{v}$ is the eigen vector and λ denotes the eigen value.
- Furthermore, suppose we calculate the covariance matrix of the data and calculates its eigen vectors then the largest eigen vector points in the direction of the largest variance of data.
- We can also say eigen values define covariance and eigen vectors define linearly independent directions.
- Thus, if we are given the point cloud of an object, by first finding the covariance of the data and then calculating its eigen vectors we can determine the orientation of the object.

2) Choice of outlier rejection for each dataset

- <u>Dataset 1:</u>



- The dataset 1 has the least number of outliers. The principle eigen vector can be seen in green and the secondary eigen vector has been denoted by blue arrow. For the first dataset the LS and LS + regularization lines have overlapped. We have also plotted the TLS line and the output of RANSAC.

ENPM 673 – HW 1

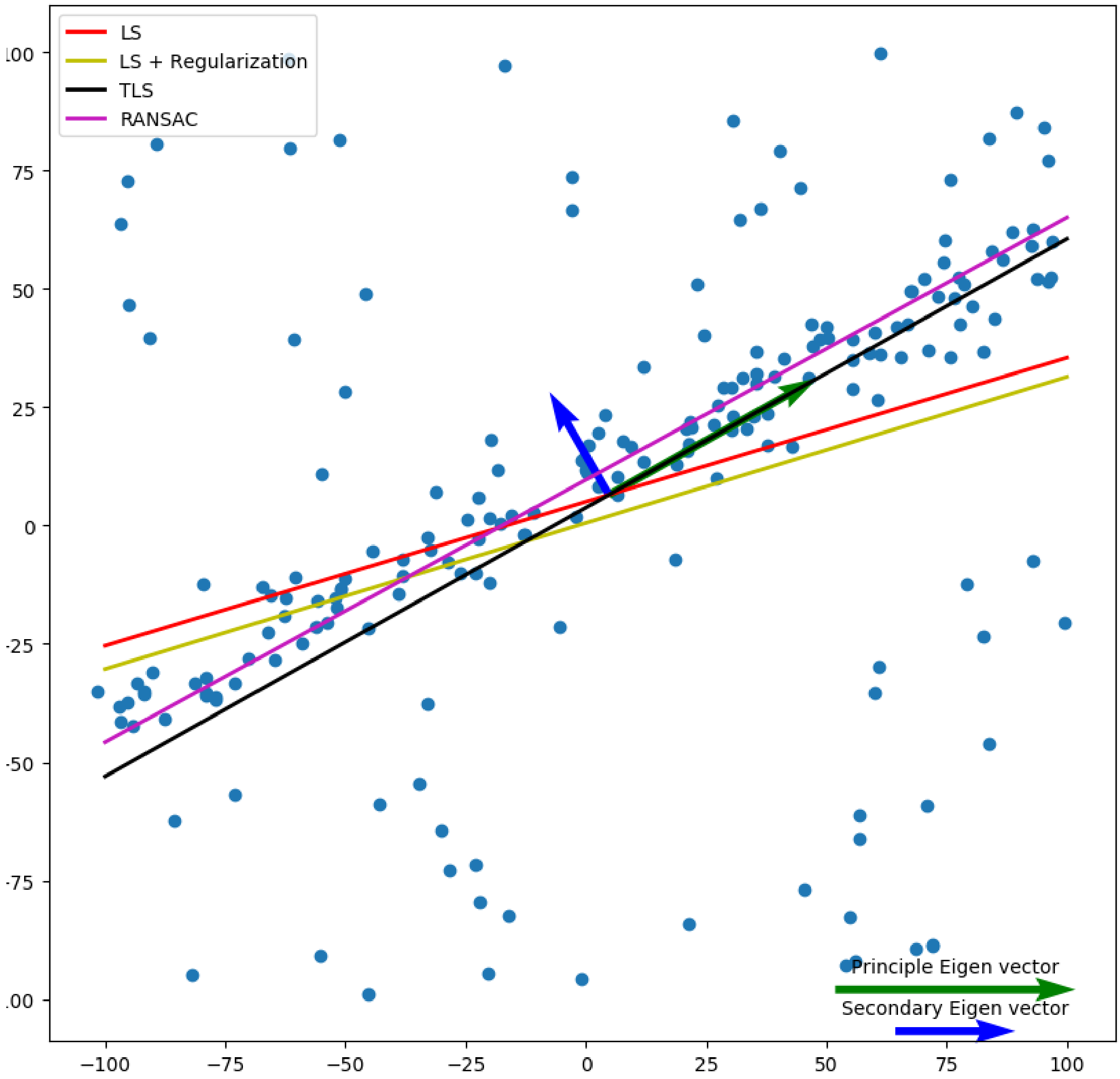
Name: Hrishikesh Tawade
UID: 116078092

Name: Krishna Bhatu
UID: 116304764

Name: Kapil Rawal
UID:116133357

- As we can see that all algorithms have performed well in this scenario. LS and LS + regularization have overlapped in this case. Although RANSAC and TLS give better output then LS and LS + regularization we can still go with using either LS or LS + regularization since they are computationally less expensive.

- Dataset 2:



- In dataset 2 the number of outliers increases, and it affects the performance of LS and LS + regularization.
- Thus, in this case we can go for TLS and avoid using RANSAC to prevent the computational expensiveness.

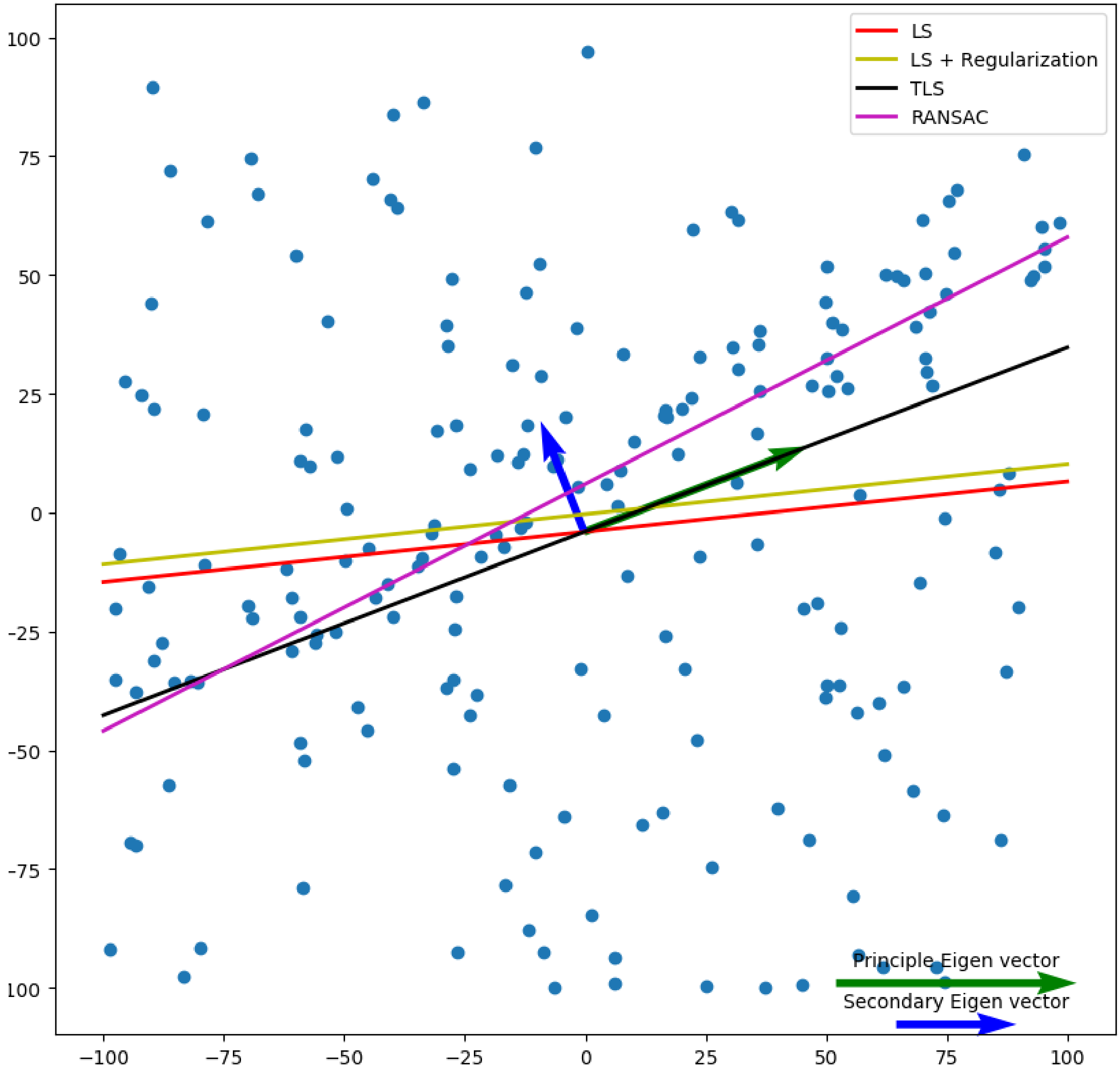
- Dataset 3
  - The third case has the maximum number of outliners and we can clearly see form the above image that LS and LS + regularization doesn't provide proper best fit line.
  - TLS also doesn't perform very well when the outlier ratio reaches between 40 to 50 percent TLS starts failing too.
  - RANSAC is the best method for this since it finds the best fit line for the maximum number of inliners for a given threshold and given iterations.
  - This it clearly rejects the outliners and gives us a proper best fit line.

ENPM 673 – HW 1

Name: Hrishikesh Tawade
UID: 116078092

Name: Krishna Bhatu
UID: 116304764

Name: Kapil Rawal
UID:116133357



3) Limitations of all the algorithms

- <u>Linear Regression</u>
  - By its nature, linear regression only looks at linear relationships between variables. Which means it assumes that there is a straight-line relationship between them. Sometimes this is incorrect. For example, the relationship between income and age is curved.
  - Outliers are data that are surprising. Outliers can be univariate (based on one variable) or multivariate. If you are looking at age and income, univariate outliers would be things like a person who is 118 years old, or one who made $12 million last year. Based on the images above we can see that outlier causes the best fitting approximation line to shift.
  - This method calculates vertical offsets which is not possible if the best fit line is vertical.

- <u>Total Linear Regression (Perpendicular offset)</u>
  - This method tries to minimize the perpendicular distance of a point from the best-fit line.
  - But the absolute value function given below does not have continuous derivatives and hence the square of the perpendicular distances is minimized instead.

$$R_\perp \equiv \sum_{i=1}^{n} \frac{|y_i - (a + b\,x_i)|}{\sqrt{1+b^2}}.$$

Name: Hrishikesh Tawade UID: 116078092
Name: Krishna Bhatu UID: 116304764
Name: Kapil Rawal UID:116133357

- The parameters obtained from minimizing the square of above equations have unwieldy form and this formulation cannot be extended for higher order polynomials. More details can be found on this LINK.
- Also, the TLS doesn't cope up well with the outliers.

- Ridge Regression or Tikhonov regularization:
  - The choice of lambda is a judgmental one. While formal methods have been developed for making this choice, these methods have their own limitations. More info is found on this LINK
  - Although, the ridge regression works well for noise removal but fails as the number of outliers goes on increasing.

- RANSAC:
  - A disadvantage of RANSAC is that there is no upper bound on the time it takes to compute these parameters (except exhaustion). When the number of iterations computed is limited the solution obtained may not be optimal, and it may not even be one that fits the data in a good way. In this way RANSAC offers a trade-off; by computing a greater number of iterations the probability of a reasonable model being produced is increased. (LINK)
  - Moreover, RANSAC is not always able to find the optimal set even for moderately contaminated sets and it usually performs badly when the number of inliers is less than 50%.
  - Another disadvantage of RANSAC is that it requires the setting of problem-specific thresholds.
  - RANSAC can only estimate one model for a particular data set. As for any one-model approach when two (or more) model instances exist, RANSAC may fail to find either one.
  - The Hough transform is one alternative robust estimation technique that may be useful when more than one model instance is present