Least Squares Report

Rishab Parthasarathy

5 April 2022

Contents

1 Introduction

This report details various least-squares fits of discrete data sets (x_i, y_i), which have been plotted as curve fits and compared to empirical calculations in Logger Pro.

These least squares fits were calculated by creating functions $F(x_i, q_1, \ldots, q_n)$ to fit the data set. Specifically, the error function

$$E = \frac{1}{2}\sum_i (y_i - F(x_i, q_1, \ldots, q_n))^2 \tag{1}$$

was minimized using gradient descent with an adaptive learning factor λ, where as presented in the lab document,

$$\begin{aligned} \Delta q_j &= -\lambda \frac{\partial E}{\partial q_j} \\ \frac{\partial E}{\partial q_j} &= -\sum_i \left[(y_i - F(x_i, q_1, \ldots, q_n)) \frac{\partial F(x_i, q_1, \ldots, q_n)}{\partial q_j} \right] \end{aligned} \tag{2}$$

These internal partial derivatives were calculated with applications of the five-point stencil, which were optimized and parallelized using NumPy matrix multiplication techniques in Python.

The various methods detailed in further parts will be compared based on the RMS error, which for N points can be computed as

$$RMSE = \sqrt{\frac{\sum_i (y_i - F(x_i, q_1, \ldots, q_n))^2}{N}} \tag{3}$$

This RMS Error is not the same error as the one presented in the lab document, but this definition is used because it standardizes the value of the error no matter the number of data points and also allows direct comparison to Logger Pro, which utilizes the RMS Error for its curve fits.

1.1 Various Notes

In this project, all code was completed in Python and all plots were generated with Matplotlib. Because of the use of Python, no issues with floating point arithmetic were detected even for arbitrarily small h in the calculation of the partial derivative via the five-point stencil. In addition, the least squares algorithm was run until either the maximum number of iterations or the error threshold was hit.

2 Linear Fit

For the linear fit, velocity-time data for constant acceleration was taken from the website (`https://plab193.wordpress.com/2018/02/14/constant-acceleration-notes/`) as shown in Table 1:

Table 1: **Time-Velocity Data**

Time (s)	Velocity (m/s)
0.485	0.704
0.594	0.918
0.686	1.089
0.767	1.236
0.840	1.367
0.907	1.486
0.970	1.597
1.028	1.700
1.084	1.798
1.137	1.890

Starting from a fit of $F(x, q_2, q_3) = q_2 x + q_3$ with $q_2 = 1$, $q_3 = 0$, the least squares converged to $q_2 = 1.800$ and $q_3 = -0.1509$ and an error threshold of $RMSE = 0.007$ within 317 iterations.

Compared to Logger Pro values of $q_2 = 1.809$ and $q_3 = -0.1585$ with $RMSE = 0.007549$, my model converged to slightly greater accuracy but still with relatively similar fit.

A plot of the finished fit with the data points is presented in Fig. 1.

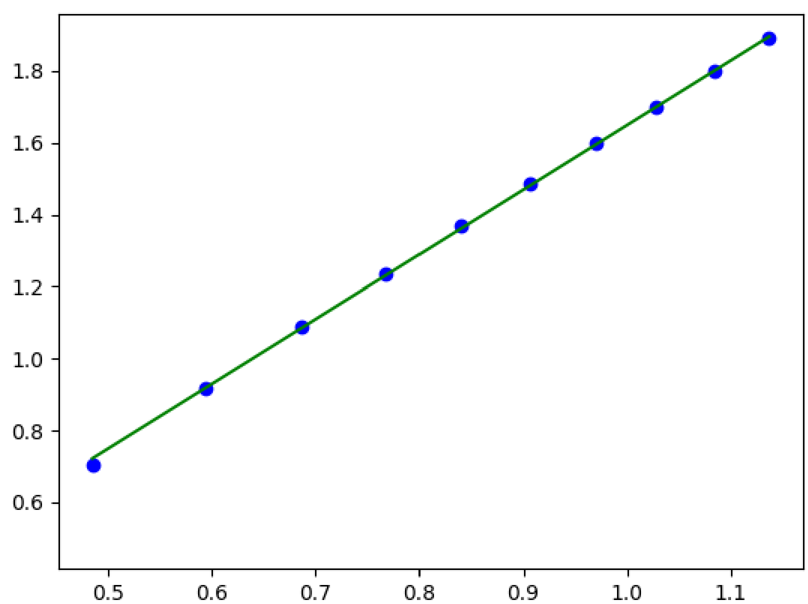

Figure 1: Figure 1 depicts the completed linear fit for the time-velocity data after hitting the error threshold of 0.007.

3 Quadratic Fit

For the quadratic fit, position-time data for constant acceleration was taken from the website (`https://plab193.wordpress.com/2018/02/14/constant-acceleration-notes/`) as shown in Table 2:

Table 2: **Position-Velocity Data**

Position (m)	Velocity (m/s)
0.343	0.1
0.485	0.2
0.594	0.3
0.686	0.4
0.767	0.5
0.840	0.6
0.907	0.7
0.970	0.8
1.028	0.9
1.084	1.0
1.137	1.1

Starting from a fit of $F(x, q_1, q_2, q_3) = q_1x^2 + q_2x + q_3$ with $q_1 = 0.5$, $q_2 = 0.5$, $q_3 = 0$, the least squares converged to $q_1 = 0.85386$, $q_2 = -0.003988$ and $q_3 = 0.00097158$ and an error of $RMSE = 0.0003145$ within 20000 iterations.

Compared to Logger Pro values of $q_1 = 0.85389$, $q_2 = -0.003988$ and $q_3 = 0.0009717$ with $RMSE = 0.0003688$, my model converged to slightly greater accuracy but still with relatively similar fit.

A plot of the finished fit with the data points is presented in Fig. 2.

4 Cubic Fit

For the cubic fit, artificial data was created based on the function $y(x) = x^3 + 2x^2 + 3x + 4$ as shown in Table 3.

Starting from a fit of $F(x, q_0, q_1, q_2, q_3) = q_0x^3 + q_1x^2 + q_2x + q_3$ with $q_0 = 0.5$, $q_1 = 0.5$, $q_2 = 0.5$, $q_3 = 0$, the least squares converged to $q_0 = 0.9937$, $q_1 = 2.0892$, $q_2 = 3.1284$ and $q_3 = 2.7647$ and an error of $RMSE = 1.2526$ within 100000 iterations.

Compared to Logger Pro values of $q_0 = 0.9992$, $q_1 = 2.055$, $q_2 = 2.970$ and $q_3 = 3.513$ with $RMSE = 1.681$, my model converged to slightly greater accuracy but still with relatively similar fit.

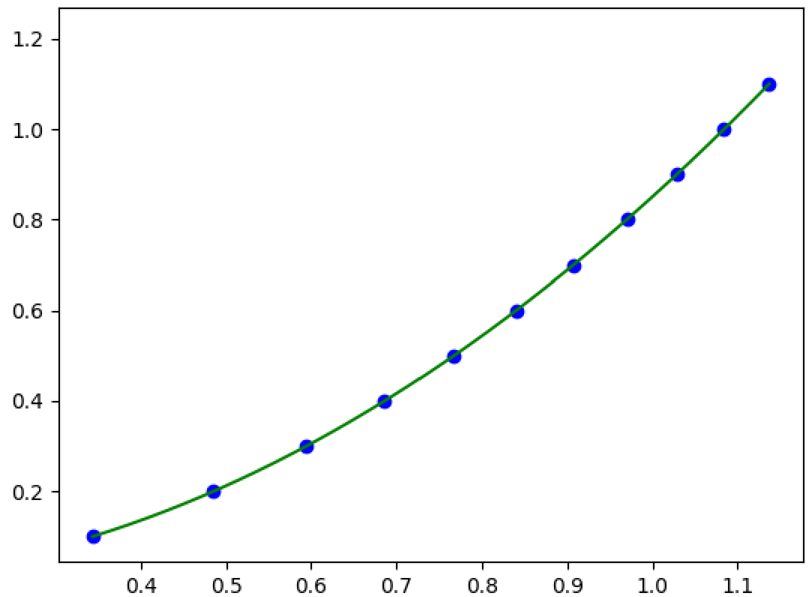

Figure 2: Figure 2 depicts the completed quadratic fit for the position-velocity data after hitting the iteration number of 20000.

Table 3: **Artificial Cubic Data**

Value of x	Value of y
-5	-85
-1	2
1	11
2	24
3	56
4	113
5	196
8	670

A plot of the finished fit with the data points is presented in Fig. 3.

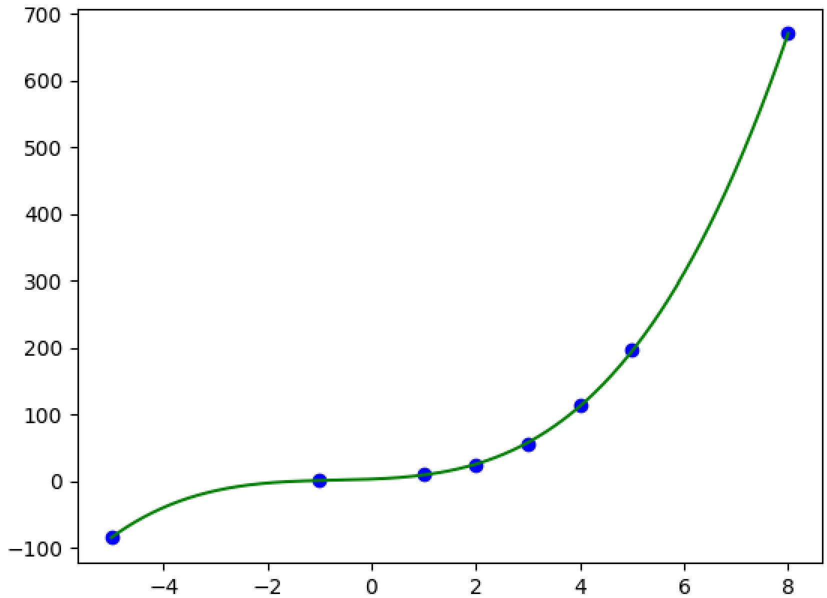

Figure 3: Figure 3 depicts the completed cubic fit for the artificial cubic data after hitting the iteration number of 100000.

5 Gaussian Fit

For the Gaussian fit, Geiger Histogram data provided by Ms. Peregrino was fit as shown in Table 4.

Starting from a fit of $F(x_i, q_0, q_1, q_2, q_3) = q_0^2 e^{-\frac{(x_i - q_1)^2}{q_2^2}} + q_3^2$ with $q_0 = 1$, $q_1 = 10$, $q_2 = 1$, $q_3 = 1$, the least squares converged to $q_0 = 531.6$, $q_1 = 12.653$, $q_2 = 6.284$ and $q_3 = 34.286$ and an error of $RMSE = 10710$ within 200000 iterations.

Compared to Logger Pro values of $q_0 = 531.8$, $q_1 = 12.654$, $q_2 = 6.292$ and $q_3 = 30.597$ with $RMSE = 11260$, my model converged to slightly greater accuracy but still with relatively similar fit.

A plot of the finished fit with the data points is presented in Fig. 4.

6 Sinusoidal Fit

For the Sinusoidal fit, the fit used was $F(x_i, q_0, q_1, q_2, q_3) = q_0 \sin{(q_1 x + q_2)} + q_3$.

Table 4: **Geiger Histogram Data**

Value of x	Value of y
0	39
1	483
2	2331
3	8094
4	20922
5	44946
6	80292
7	125231
8	174923
9	220993
10	255802
11	278362
12	283497
13	276316
14	257485
15	231036
16	201821
17	170357
18	139010
19	111864
20	87295
21	65983
22	48481
23	34820
24	24497
25	16187
26	10714
27	6968
28	4178
29	2546
30	1512
31	846
32	451
33	244
34	146
35	61
36	35
37	20
38	5
39	5
40	0
41	1

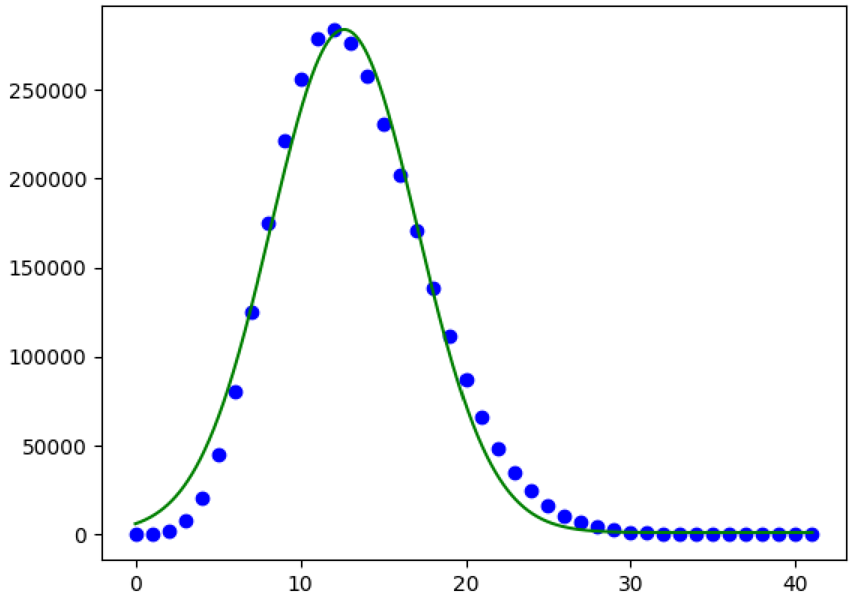

Figure 4: Figure 4 depicts the completed Gaussian fit for the Geiger histogram data after hitting the iteration number of 200000.

6.1 Dataset 1

The first dataset used was contained within 'UCrBHe1.txt' as provided by Ms. Peregrino and described in Table 5.

Starting from a fit of $F(x_i, q_0, q_1, q_2, q_3) = q_0 \sin(q_1 x + q_2) + q_3$ with $q_0 = 1$, $q_1 = 1$, $q_2 = 1$, $q_3 = 0$, the least squares converged to $q_0 = 73.1613$, $q_1 = 6.2624$, $q_2 = -3.1411$, $q_3 = -6.633$, and an error of $RMSE = 2.6496$ within 50000 iterations.

Compared to Logger Pro values of $q_0 = 73.1615$, $q_1 = 6.2624$, $q_2 = 3.1420$ and $q_3 = -6.633$ with $RMSE = 3.094$, my model converged to slightly greater accuracy but still with relatively similar fit, with variation in q_2 covered by the period of 2π.

A plot of the finished fit with the data points is presented in Fig. 5.

6.2 Dataset 2

The second dataset used was contained within 'UCrBMg2.txt' as provided by Ms. Peregrino and described in Table 6.

Starting from a fit of $F(x_i, q_0, q_1, q_2, q_3) = q_0 \sin(q_1 x + q_2) + q_3$ with $q_0 = 1$, $q_1 = 1$, $q_2 = 1$, $q_3 = 0$, the least squares converged to $q_0 = 101.792$, $q_1 = 6.2727$, $q_2 = -0.007465$, $q_3 = -21.136$, and an

Table 5: **Sinusoidal Dataset 1**

Value of x	Value of y
0	-6.694
0.05	-26.776
0.1	-53.551
0.2	-74.971
0.25	-80.327
0.3	-74.302
0.4	-50.204
0.45	-30.122
0.5	-3.347
0.55	8.033
0.7	65.6
0.75	66.939
0.8	63.592
0.9	36.816
0.95	16.735

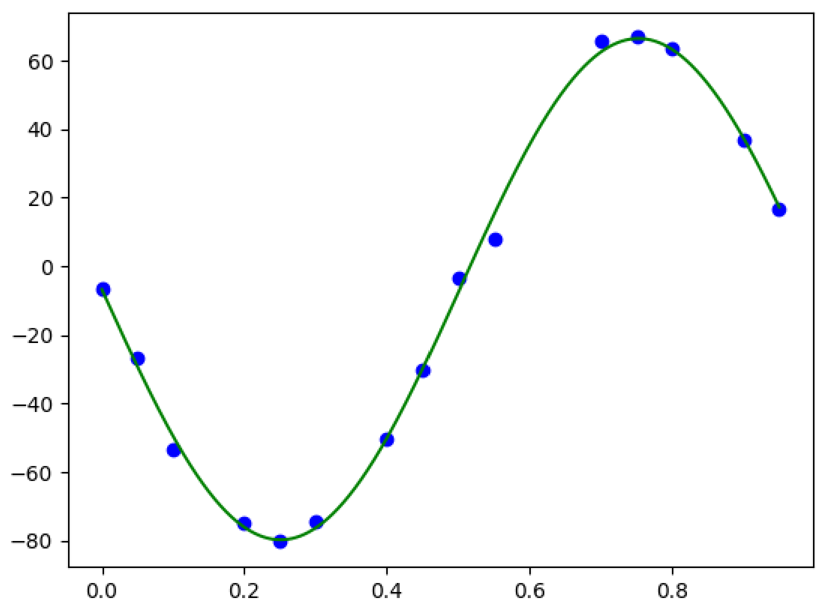

Figure 5: Figure 5 depicts the completed sinusoidal fit for the 'UCrBHe1.txt' dataset after hitting the iteration number of 50000.

Table 6: **Sinusoidal Dataset 2**

Value of x	Value of y
0	-20.036
0.05	-1.336
0.1	43.411
0.2	80.144
0.25	86.822
0.3	66.786
0.4	40.072
0.45	10.018
0.5	-20.036
0.55	-53.429
0.7	-110.865
0.75	-123.555
0.8	-121.551
0.9	-84.151
0.95	-53.429

error of $RMSE = 4.905$ within 50000 iterations.

Compared to Logger Pro values of $q_0 = 101.864$, $q_1 = 6.2733$, $q_2 = 6.2755$ and $q_3 = -21.361$ with $RMSE = 5.728$, my model converged to slightly greater accuracy but still with relatively similar fit, with variation in q_2 covered by the period of 2π.

A plot of the finished fit with the data points is presented in Fig. 6.

7 Conclusion

Overall, throughout the test of the linear, quadratic, cubic, Gaussian, and sinusoidal fits, my gradient-based least-squares approach achieved comparable results and good fits compared to empirical Logger Pro values.

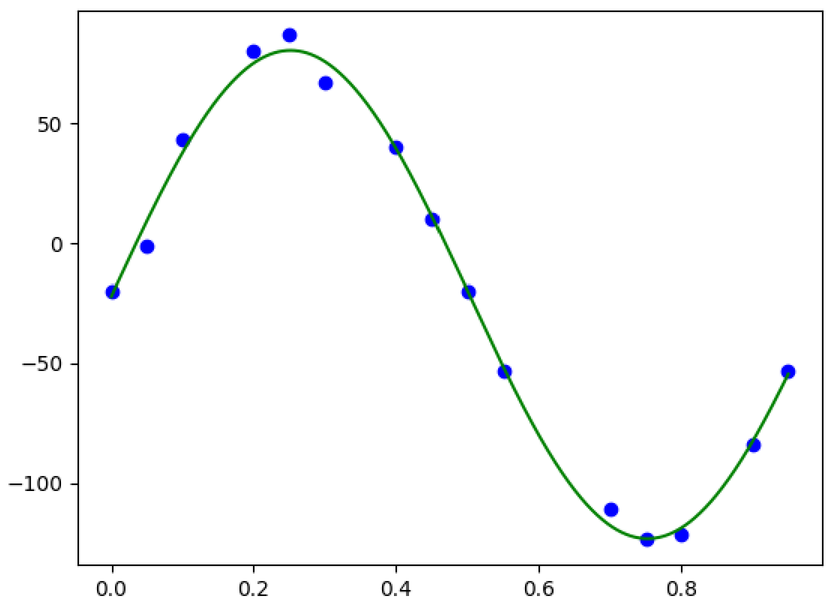

Figure 6: Figure 6 depicts the completed sinusoidal fit for the 'UCrBMg2.txt' dataset after hitting the iteration number of 50000.