Uncertainty analysis of linear least square fitting apply to non-linear model

T Arunthong, S Thuenhan, P Wongsripan, S Chomkokard, W Wongkokua and N Jinuntuya

Department of Physics, Faculty of Science, Kasetsart University, Bangkok, 10900, Thailand

fscinpr@ku.ac.th

Abstract. Linear least square is a common method to fit two-dimensional data set with linear relation. In non-linear case we need to do some linearization before fitting. Parameters can be extracted from the fitted line and used to predict new dependent variable values at new independent variable points. In this report we investigate the uncertainty introduced to a non-linear model prediction from linear least square fitting. As an example, we fit a data set with the relation $D = a/h$, which is the physical model of measurement we are interested. The fitting goes well, and we get the curve with R-square very closed to one. However, when we use the fitted parameters to actual measurements, the accuracy is poor, especially at large y. At first, we expect that this due to the functional form of the model. When compare with high order polynomial fitting we realize that this is not the case, since, for example, the sixth-degree polynomial gives less than one percent error, about 10 times less than linear fitting prediction. It is because the linearization and the inverse transformation to the original space in the linear least square method that give rise to the uncertainty. Our analysis can be generalized to any non-linear model prediction. We expect our results to be a caution to anyone using linear fitting to their non-linear model prediction.

1. Introduction

Curve fitting is a common task in data analysis. In this work we analyze data of distance measurements from a parallax laser range finder [1]. The optical triangulation is used for range finding, as shown in figure 1. The target distance is calculated with simple trigonometry to be

$$D = Hd/h. \tag{1}$$

Here the distance H and d are fixed as the system parameters. The target distance D is inversely proportional to the distance of laser point from image center h. We can use mathematical relation of the form

$$D = a/h, \tag{2}$$

to model our measurement. As a standard procedure, we perform an experiment by measure h at various known values of D to produce a calibration curve. From the curve we can estimate the system parameter $a = Hd$. This parameter will be used to predict the target distance in actual measurement.
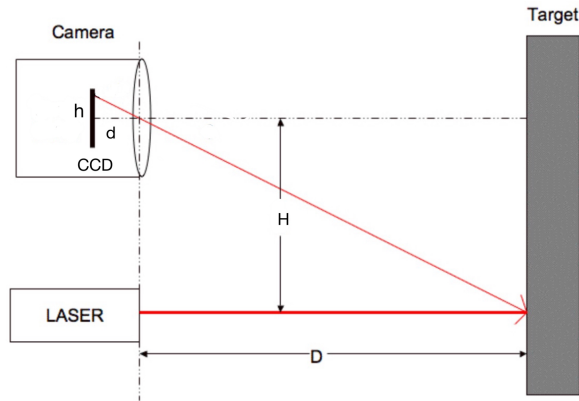
Figure 1. Configuration of a parallax laser range finder.

2. Linear least square fitting

Now curve fitting plays its role in our work. We need to fit (h, D) data set to estimate a. Linear least square is the simplest method to do the job. Since the model in equation (2) is non-linear, we need to do linearization before applying linear least square fitting. It is straightforward to follow the physical model by fitting $1/D$ versus h to obtain $\alpha = 1/a$ using linear least square method. Table 1 shows our experimental results. The plot of D versus h is shown in figure 2. Figure 3 shows linear relation between $1/D$ versus h, and the corresponding fitted parameter and R-square are shown in table 2. We can see that the calibration line fits the experimental points quite well.

Table 1. Experimental results of the position of laser point h and target distance D.

h (pixel)	D (m)	$E = 1/D$ (m^{-1})	h (pixel)	D (m)	$E = 1/D$ (m^{-1})
1108	0.535	1.87	321	1.818	0.550
1017	0.583	1.72	215	2.755	0.363
879	0.668	1.50	191	3.110	0.322
751	0.778	1.29	167	3.617	0.276
686	0.850	1.18	151	4.048	0.247
583	0.996	1.00	131	4.626	0.216
408	1.425	0.702			
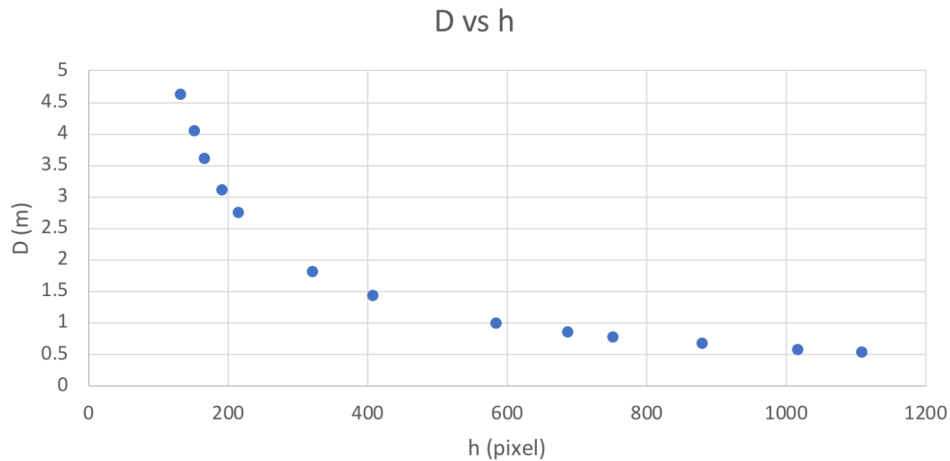
Figure 2. The plot of D versus h from data in table 1.
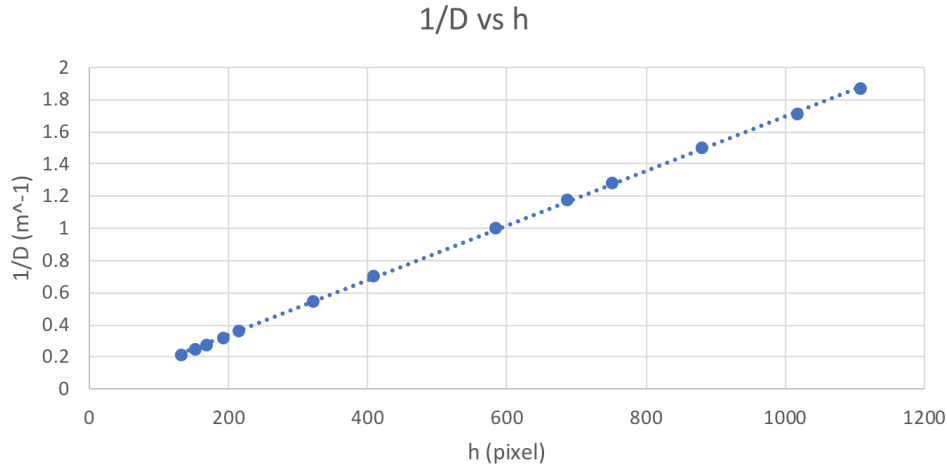
Figure 3. Linear relation between $1/D$ versus h.

Table 2. Slope α, y-interception and R-square from figure 3.

	α (m^{-1})	y-intercept (m^{-1})	R^2
Value	0.001699	0.000	0.9998
Uncertainty	0.000008	0.005	-

We next test our system by performing another data set, as shown in table 3. We compute percentage difference between our predicted results and the distances measured from a commercial laser distance meter. It is obvious that the percentage difference increases with the target distance. At first, we expect that this due to the functional form of the model. To test our hypothesis, we perform non-linear fittings to the first data set using the sixth-order polynomial fittings. Percentage differences between the polynomial predictions and the laser distance meter are also shown in table 3. The percentage difference decreases with increasing target distance. Functional form of our model may not be the cause of discrepancies from the prediction of linear least square.

Table 3. Comparison of target distance between a commercial laser distance meter and the models predictions.

Distance meter (m)	Linear prediction (m)	%diff	Polynomial prediction (m)	%diff
1.160	1.177	1.47	1.196	3.10
1.585	1.608	1.44	1.588	0.192
3.357	3.303	1.62	3.380	0.675
3.846	3.746	2.65	3.864	0.475
4.365	4.244	2.81	4.377	0.268

3. Uncertainty analysis

We try to understand the above results in term of propagation of uncertainty [2, 3]. As mention above, we need to linearize the data before doing linear least square fitting. In the process to obtain the calibration line, we transform D to $E = 1/D$. The value of $\alpha = 1/a$ and its uncertainty are evaluated using standard procedure. We use here the Excel LINEST command to do the calculations, which the results are shown in table 2. Since the relation between E and h is linear, uncertainty $u(E)$ of the predicted E is independent of both E and h. When we convert E back to D, the uncertainty of D is

$$u(D) = u(E)/E^2 = D^2 u(E).$$ (3)

It is obvious that $u(D)$ is increasing with D in this case.

To answer the question why percentage error is decreasing with increasing D in the polynomial fit, let examine the propagation of uncertainty. For the polynomial fit the predicted D is obtained from

$$D = c_0 + c_1 h + c_2 h^2 + ... + c_n h^n,$$ (4)

where n is the order of polynomial. The uncertainty of D is

$$u(D) = u(c_0) + |h|u(c_1) + |h|^2 u(c_2) + ... + |h|^n u(c_n) + |c_1 + 2c_2 h + ... + nc_n h^{n-1}|u(h)$$ (5)

In this case we can see that $u(D)$ is increasing with h. Since $D = a/h$, it means that $u(D)$ is decreasing with increasing D.

4. Conclusion

To conclude, by examine the uncertainty propagation in linear least square and polynomial fitting applied to our simple non-linear model, we find that the behavior of uncertainty depends strongly on the fitting method. Linear least square fitting gives the calibrated parameter that directly related to the physical model but suffer from large percentage error at large distance. Polynomial fitting gives smaller error at large distance. However, parameters from polynomial fitting have no physical meaning. It also suffers from oscillation when the number of data are small. Using both methods should give better results, with some trade-off have to be optimized.

References
[1] *Parallax Laser Range Finder* (#28044) data sheet (https://www.parallax.com/: Access 30/04/2018) Paralax Inc
[2] Taylor J R 1997 *An Introduction to Error Analysis 2nd ed.* (University Science Books: USA)
[3] Hughes I G and Hase T P A 2010 *Measurement and their Uncertainties* (New York: Oxford University Press)