Parameter estimation in exponential models by linear and nonlinear fitting methods[*]

Ping YANG[†1], Chao-peng WU[2], Yi-lu GUO[2], Hong-bo LIU[2], Hui HUANG[†‡2], Hang-zhou WANG[2],

Shu-yue ZHAN[2], Bang-yi TAO[3], Quan-quan MU[4], Qiang WANG[1], Hong SONG[2]

([1]*School of Digital Media & Design, Hangzhou Dianzi University, Hangzhou 310018, China*)

([2]*Ocean College, Zhejiang University, Zhoushan 316021, China*)

([3]*State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography,*

State Oceanic Administration, Hangzhou 310012, China)

([4]*State Key Laboratory of Applied Optics, Changchun Institute of Optics, Fine Mechanics and Physics,*

Chinese Academy of Sciences, Changchun 130033, China)

[†]E-mail: yangping@hdu.edu.cn; huih@zju.edu.cn

Received Nov. 3, 2016; Revision accepted Feb. 16, 2017; Crosschecked Feb. 28, 2017

Abstract: Estimation of unknown parameters in exponential models by linear and nonlinear fitting methods is discussed. Based on the extreme value theorem and Taylor series expansion, it is proved theoretically that the parameters estimated by the linear fitting method alone cannot minimize the sum of the squared residual errors in the measurement data when measurement noise is involved in the data. Numerical simulation is performed to compare the performance of the linear and nonlinear fitting methods. Simulation results show that the linear method can obtain only a suboptimal estimate of the unknown parameters and that the nonlinear method gives more accurate results. Application of the fitting methods is demonstrated where the water spectral attenuation coefficient is estimated from underwater images and imaging distances, which supports the improvement in the accuracy of parameter estimation by the nonlinear fitting method.

Key words: Exponential model; Parameter estimation; Linear least squares; Nonlinear fitting
http://dx.doi.org/10.1631/FITEE.1601683   **CLC number:** O433.1

1 Introduction

There are many physical models based on the exponential function. For example, when light penetrates water, its intensity decreases exponentially with underwater distance (Wei and Lee, 2013; Adi, 2015); the transverse amplitude distribution of the fundamental-mode radiation field generated by a laser resonator generally reflects the Gaussian function

(Liu *et al.*, 2013); the central pressure difference of tropical cyclones after making landfall may decay exponentially with time (Vickery, 2005); and the activity of a radioactive sample in a quasi-stationary state obeys the exponential decay law (Novković *et al.*, 2006; Semkow, 2007).

To determine the unknown parameters of the exponential model (especially the unknown parameter in the exponential term, e.g., the spectrum attenuation coefficient of a water column), it is common to fit the parameterized model to the measurement data. In general, this involves taking the sum of squared errors between the measured and fitted values as an objective function, and applying numerical optimization algorithms (Hartley, 1961; Peterson *et al.*, 2010) to optimize the estimates of the unknown parameters,

such that the objective function is minimized and the estimates of the parameters are as close to the true values as possible.

Alternatively, the exponential model can be transformed into a linear one by taking a logarithmic transformation on the exponential term (Lieb, 1997; Kaeli *et al.*, 2011; He and Li, 2015; Bardaji *et al.*, 2016; Simon and Shanmugam, 2016). After linearization of the exponential model, the linear least squares (LLS) method is applied to estimate the unknown parameters. This method is referred to as the linear fitting method in this paper; this method has obvious advantages, such as simplifying and reducing the amount of calculation, and has therefore been widely adopted.

However, the estimates of parameters achieved by the linear fitting method are not optimal in the sense that there is excessive residual error between the measurement data and the model when noise exists in the measurement data. The discrepancy between the true values of the parameters and the estimates can be further minimized. To verify this statement, we proceed from the expression of the sum of the squared residual errors and take the Taylor expansion and extreme value theorem as a theoretical basis. Numerical simulation further indicates that the nonlinear fitting method can better minimize the sum of squared errors and optimize the estimates of unknown parameters.

The contribution of this paper lies mainly in the theoretical analysis and numerical simulation, which shows that linearization of the exponential model followed by linear least squares fitting does not give optimal estimates of the unknown parameters when noise is involved in the measurement data. A combination of the linear method and the optimization algorithm leads to more accurate results.

2 Theoretical analysis

2.1 Problem formulation

A basic exponential model can be written as

$$y = f(x) = a_0 e^{b_0 x} + p \cdot n, \qquad (1)$$

where x is the independent variable, y is the dependent variable, and f represents the mapping from x to y.

Both x and y are known or can be measured. The term $p \cdot n$ defines the measurement noise (Mascaro *et al.*, 2011; Lai *et al.*, 2013), where n represents noise with a certain distribution (e.g., Gaussian distribution), and p is a coefficient characterizing the magnitude of the noise. Parameters a_0 and b_0 are both unknown in the model and are to be estimated by fitting a set of m measurement data points $\{x_i, y_i\}_{i=1}^m$.

The estimates of parameters a_0 and b_0 are referred to as a and b, respectively. The output of the model is referred to as $\hat{y}_i$, i.e., $\hat{y}_i = a e^{bx_i}$, with i=1, 2, ..., m. The residual error Δy_i is the difference between the measurement and its estimate, i.e.,

$$\Delta y_i = y_i - \hat{y}_i. \qquad (2)$$

During the fitting of data, an objective function (denoted by S_N) is defined by the sum of the squared residual errors as

$$S_N = \sum_{i=1}^m (\hat{y}_i - y_i)^2 = \sum_{i=1}^m (a e^{bx_i} - y_i)^2. \qquad (3)$$

The goal of data fitting is to choose the optimal parameters a and b such that the objective function S_N is minimized, i.e.,

$$a, b = \arg \min_{a^*, b^*} \sum_{i=1}^m (a^* e^{b^* x_i} - y_i)^2. \qquad (4)$$

2.2 Linear fitting method

Ignoring the measurement noise and performing logarithmic transformation (Xiao *et al.*, 2011) on both sides of Eq. (1), we have

$$\ln y = b_0 x + \ln a_0. \qquad (5)$$

Denoting $z = \ln y$, we can obtain the data set $\{x_i, z_i\}_{i=1}^m$ from the measurement data $\{x_i, y_i\}_{i=1}^m$ with $z_i = \ln y_i$. Denoting $\theta_0 = \ln a_0$, Eq. (5) is changed to a linear equation:

$$z = b_0 x + \theta_0. \qquad (6)$$

Because Eq. (6) is linear, parameters θ_0 and b_0 can be estimated by the LLS method as follows.

Denote θ as the estimate of θ_0. The model estimate can be expressed as $\hat{z}_i = bx_i + \theta$, where $\hat{z}_i$ is the estimate of z_i. The sum of squared errors (denoted as S) between estimate $\hat{z}_i$ and measurement z_i can be written as

$$S = \sum_{i=1}^{m} (\hat{z}_i - z_i)^2 = \sum_{i=1}^{m} (bx_i + \theta - z_i)^2. \tag{7}$$

To minimize S, the estimated parameters θ and b should satisfy the following conditions:

$$\begin{cases} \dfrac{\partial S}{\partial b} = 2\sum_{i=1}^{m} x_i (bx_i + \theta - z_i) = 0, \\ \dfrac{\partial S}{\partial \theta} = 2\sum_{i=1}^{m} (bx_i + \theta - z_i) = 0. \end{cases} \tag{8}$$

Because the conditions in Eq. (8) are linear equations of unknowns θ and b, Eq. (8) can be expressed in matrix form as

$$\underbrace{\begin{bmatrix} \sum_{i=1}^{m} x_i^2 & \sum_{i=1}^{m} x_i \\ \sum_{i=1}^{m} x_i & m \end{bmatrix}}_{A} \cdot \underbrace{\begin{bmatrix} b \\ \theta \end{bmatrix}}_{X} = \underbrace{\begin{bmatrix} \sum_{i=1}^{m} x_i z_i \\ \sum_{i=1}^{m} z_i \end{bmatrix}}_{B}. \tag{9}$$

When the coefficient matrix $A \in \mathbb{R}^{2 \times 2}$ has full column rank, we have the solution:

$$X = A^{-1} B, \tag{10}$$

which is the LLS solution to the problem of exponential model fitting. The solution obtained by linear fitting is denoted as θ_L and b_L, and thus we have $a_L = e^{\theta_L}$.

2.3 Nonlinear fitting method

The objective function S_N is a continuous function that is dependent on parameters a and b. According to the extreme value theorem, the estimates a and b should satisfy the following conditions when S_N reaches its minimum:

$$\begin{cases} \dfrac{\partial S_N}{\partial a} = 2\sum_{i=1}^{m} (ae^{bx_i} - y_i)e^{bx_i} = 0, \\ \dfrac{\partial S_N}{\partial b} = 2\sum_{i=1}^{m} ax_i (ae^{bx_i} - y_i)e^{bx_i} = 0. \end{cases} \tag{11}$$

Because the parameters in Eq. (11) are not linear, it is difficult to obtain analytical expressions of a and b; however, numerical solutions can be obtained by numerical optimization algorithms. Proceeding from the initial parameter estimates, we search for better estimates of the parameter to minimize S_N until the algorithm converges. Commonly used optimization algorithms include the steepest descent method, Levenberg-Marquardt (LM) algorithm, and Gauss-Newton algorithm (Levenberg, 1944; Marquardt, 1963; Floudas and Pardalos, 2001).

In the steepest descent method, the objective function is iteratively reduced by updating the estimate of the parameter in the direction opposite to the gradient of the objective function. That is,

$$\boldsymbol{g}_{i+1} = \boldsymbol{g}_i - t_i \nabla S(\boldsymbol{g}_i), \tag{12}$$

where $i = 0, 1, 2, \ldots$, $\boldsymbol{g}_i$ is the parameter estimated in the ith iteration, $t_i > 0$ is the step factor, and $\nabla S(\boldsymbol{g}_i)$ is the gradient of the objective function S.

The Gauss-Newton algorithm is a modification of Newton's method for minimizing the sum of squared function values. The advantage of the Gauss-Newton method is that the second derivatives of the objective function, which can be challenging to compute, are not required. The estimate of the unknown parameters is updated as

$$\boldsymbol{g}_{i+1} = \boldsymbol{g}_i + (J_f^\mathrm{T} J_f)^{-1} J_f^\mathrm{T} \cdot \boldsymbol{r}(\boldsymbol{g}_i), \tag{13}$$

where J_f is the Jacobian of the function f (i.e., the exponential function in our case), and 'T' denotes the transpose of the matrix. Vector $\boldsymbol{r} = [\Delta y_1, \Delta y_2, \ldots, \Delta y_m]^\mathrm{T}$ is composed of the residual errors of each measurement datum as in Eq. (2).

The LM algorithm is a combination of the steepest descent method and the Gauss-Newton method. It is widely adopted in various nonlinear minimization problems. The LM algorithm behaves like a steepest descent method and increases the step factor when the current estimates are far from their

true values. When the objective function is reduced, the step factor is decreased and the algorithm is more like the Gauss-Newton method.

Because the objective function may have multiple local minima, all these methods can converge to a local minimum, which is not necessarily the global minimum of the objective function. Therefore, the choice of the initial estimate is important during the optimization. Usually the initial values can be chosen from physical insight, prior knowledge, etc. Because the linear method also provides an estimate of the unknown parameter (i.e., a_L and b_L), a_L and b_L are then used as the initial guess in the optimization algorithm. The combination of the linear fitting method and the optimization algorithm is called the nonlinear fitting method in this paper.

2.4 Influence of noise on fitting methods

2.4.1 Optimality of estimates a_L and b_L

Although θ_L and b_L are LLS solutions of Eq. (8) in the sense that the sum of squared logarithmic errors (i.e., S) is minimized, θ_L and b_L fail to minimize the sum of squared errors (i.e., S_N) in Eq. (3) when noise is involved in the measurement data. To explain this explicitly, detailed analysis is presented as follows.

In the linear fitting method, we have $y_i = e^{z_i}$ and $\hat{y}_i = e^{\hat{z}_i}$, and thus the sum of squared errors can be written as

$$S_L = \sum_{i=1}^{m} (\hat{y}_i - y_i)^2 = \sum_{i=1}^{m} (e^{\hat{z}_i} - e^{z_i})^2. \quad (14)$$

Eqs. (3) and (14) are both sums of squared errors between exponential model estimates $\hat{y}_i$ and measurements y_i. The subscript 'L' is used to emphasize that S_L is calculated from the results by the linear fitting method.

According to the extreme value theorem, to minimize S_L, the estimates θ and b should satisfy the following conditions (note that $\hat{z}_i = bx_i + \theta$):

$$\begin{cases} \dfrac{\partial S_L}{\partial b} = 2\sum_{i=1}^{m} x_i (e^{2\hat{z}_i} - e^{z_i+\hat{z}_i}) = 0, \\ \dfrac{\partial S_L}{\partial \theta} = 2\sum_{i=1}^{m} (e^{2\hat{z}_i} - e^{z_i+\hat{z}_i}) = 0. \end{cases} \quad (15)$$

Taking Taylor expansions of $e^{2\hat{z}_i}$ and $e^{z_i+\hat{z}_i}$, we have

$$e^{2\hat{z}_i} = \sum_{n=0}^{\infty} \frac{(2\hat{z}_i)^n}{n!}, \quad e^{z_i+\hat{z}_i} = \sum_{n=0}^{\infty} \frac{(z_i+\hat{z}_i)^n}{n!}. \quad (16)$$

By defining a new variable as

$$q_{i,n} = (2\hat{z}_i)^n - (z_i+\hat{z}_i)^n, \quad (17)$$

Eq. (15) is changed to

$$\begin{cases} 2\sum_{i=1}^{m}\sum_{n=0}^{\infty} x_i \dfrac{q_{i,n}}{n!} = 0, \\ 2\sum_{i=1}^{m}\sum_{n=0}^{\infty} \dfrac{q_{i,n}}{n!} = 0. \end{cases} \quad (18)$$

Expanding Eq. (18) for $n=0$ and $n=1$, we have

$$\begin{cases} 2\sum_{i=1}^{m} x_i (bx_i + \theta - z_i)+2\sum_{i=1}^{m}\sum_{n=2}^{\infty} x_i \dfrac{q_{i,n}}{n!} = 0, \\ 2\sum_{i=1}^{m} (bx_i + \theta - z_i)+2\sum_{i=1}^{m}\sum_{n=2}^{\infty} \dfrac{q_{i,n}}{n!} = 0. \end{cases} \quad (19)$$

Because the LLS solutions θ_L and b_L satisfy Eq. (8), the condition in Eq. (19) is simplified to

$$\begin{cases} 2\sum_{i=1}^{m}\sum_{n=2}^{\infty} \dfrac{x_i q_{i,n}}{n!} = 0, \\ 2\sum_{i=1}^{m}\sum_{n=2}^{\infty} \dfrac{q_{i,n}}{n!} = 0. \end{cases} \quad (20)$$

Defining the residual error as $e_i = z_i - \hat{z}_i$, the term $q_{i,n}$ can be expressed as

$$q_{i,n} = (2bx_i + 2\theta)^n - (2bx_i + 2\theta + e_i)^n. \quad (21)$$

If the residual error $e_i=0$, then we have $q_{i,n}=0$, so θ_L and b_L satisfy Eq. (20) and in turn Eq. (15). In other words, a_L and b_L obtained by the linear fitting method are just the true values of the unknown parameters a_0 and b_0, i.e., $a_L=a_0$ and $b_L=b_0$. The sum of squared errors S_L in Eq. (14) is zero.

However, in practice, noise always exists in the measurement data (i.e., the term $p \cdot n$ in Eq. (1) is

nonzero) and residual errors also exist in the fitting process, i.e., $e_i \neq 0$ and $q_{i,n} \neq 0$. Because the left side of Eq. (20) is a nonlinear infinite sum formula, it cannot be guaranteed that the condition in Eq. (20) or Eq. (15) is always satisfied. Hence, θ_L and b_L minimize only the sum of squared logarithmic errors (i.e., S in Eq. (7)) but not the original errors (i.e., S_L in Eq. (14)). Therefore, a_L and b_L are not optimal estimates of the unknown parameters a_0 and b_0.

2.4.2 Estimation error

Referring to Eq. (6), for a set of logarithmic data $\{x_i, z_i\}_{i=1}^{m}$, we have an equation set in matrix form as

$$\underbrace{\begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_m \end{bmatrix}}_{Z} = \underbrace{\begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_m & 1 \end{bmatrix}}_{C} \cdot \underbrace{\begin{bmatrix} b \\ \theta \end{bmatrix}}_{X}, \tag{22}$$

where θ and b are to be estimated. When matrix C has full column rank, the LLS solution is given by

$$X = (C^T C)^{-1} C^T Z. \tag{23}$$

Considering that noise is involved in the measurement, z_i is disturbed as

$$z_i = b_0 x + \theta_0 + \delta_{z_i}, \tag{24}$$

where δ_{z_i} is the disturbance caused by the noise. For a set of logarithmic data $\{x_i, z_i\}_{i=1}^{m}$, Eq. (24) can be expressed as

$$\underbrace{\begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_m \end{bmatrix}}_{Z} = \underbrace{\begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_m & 1 \end{bmatrix}}_{C} \cdot \underbrace{\begin{bmatrix} b_0 \\ \theta_0 \end{bmatrix}}_{X_0} + \underbrace{\begin{bmatrix} \delta_{z_1} \\ \delta_{z_2} \\ \vdots \\ \delta_{z_m} \end{bmatrix}}_{\delta_Z}. \tag{25}$$

Substituting Eq. (25) into Eq. (23), we have

$$X = X_0 + \underbrace{(C^T C)^{-1} C^T \delta_Z}_{\Delta X}, \tag{26}$$

where ΔX is the estimation error that originates from the noise. In general, the error ΔX increases with δ_Z, but is also influenced by projection matrix $(C^T C)^{-1} C^T$.

Concerning the nonlinear method, because the optimization is achieved iteratively and there is no analytical solution on the estimates, it is difficult to determine the optimality of the estimates or the estimation error theoretically. Numerical simulation is performed in Section 3 to evaluate the accuracy of the estimates.

3 Numerical simulation

To verify the theoretical analysis in Section 2, numerical simulation has been performed for data fitting. The performance of two methods has been compared as well.

3.1 Numerical model

The numerical model for data generation is set as

$$y_i = a_0 e^{b_0 x_i} + p \cdot n_i. \tag{27}$$

The distribution of the noise sequence n_i conforms to the standard Gaussian distribution, with zero mean and variance of 1. The length of the sequence (i.e., m) is set to $m=60$ and x_i is defined as $x_i=i/20$ with $i=1, 2, \ldots, 60$.

Parameters a_0 and b_0 are estimated from $\{x_i, y_i\}_{i=1}^{m}$ by the linear and nonlinear methods. During nonlinear optimization, the LM algorithm is chosen because it shows good robustness and has been used in many applications. The parameter estimated from the linear method is used as the initial guess of nonlinear fitting, to further improve the robustness of nonlinear optimization.

Because the estimation result depends on parameters a_0, b_0, and p, the values of a_0, b_0, and p are changed individually to see how the estimation errors are affected by the parameters.

The relative error between the estimates and the true value is defined as

$$\begin{cases} \varepsilon_a = \left| \dfrac{a - a_0}{a_0} \right| \times 100\%, \\[3mm] \varepsilon_b = \left| \dfrac{b - b_0}{b_0} \right| \times 100\%, \end{cases} \tag{28}$$

where ε_a and ε_b are the relative errors of estimated parameters a_0 and b_0, respectively.

The relative sum of squared errors is defined as

$$\varepsilon_{\mathrm{L}} = S_{\mathrm{L}} \bigg/ \sum_{i=1}^{m} y_i^2 \times 100\% \tag{29}$$

for linear fitting and

$$\varepsilon_{\mathrm{N}} = S_N \bigg/ \sum_{i=1}^{m} y_i^2 \times 100\% \tag{30}$$

for nonlinear fitting. In Section 3.2, ε_{L} and ε_{N} are collectively termed ε_{S}.

For a set of selected values of a_0, b_0, p, and x_i, 1000 sets of Gaussian random sequences n_i are generated and added to y_i to perform 1000 repeated numerical trials. After 1000 trials, the mean values of ε_a, ε_b, ε_{L}, and ε_{N} are evaluated and compared.

3.2 Numerical results

3.2.1 Effect of parameter a_0

To observe how the errors change with a_0, parameters b_0 and p are set as $b_0=-1$ and $p=0.01$, and a_0 changes from -2 to 2.

Fig. 1 shows the relative estimation error with respect to a_0. For example, in the linear method, ε_a reaches its maximum when a_0 is around zero, but ε_a declines from 25.7% to 0.1% and ε_b declines from 33.0% to 0.1% as a_0 increases from 0 to 2. Fitting accuracy is improved as a_0 increases, because the influence of noise in the measurement is reduced; i.e., the signal-to-noise ratio (SNR) is improved. Compared with the linear method, the nonlinear method provides higher accuracy, with the estimation error one order of magnitude smaller. Apart from that, the result of the nonlinear method is less influenced by the magnitude of a_0, showing good robustness of the method.

As shown in Fig. 2, both ε_{L} and ε_{N} decrease as the absolute value of a_0 increases, but ε_{L} is always 1.2–2 times greater than ε_{N}.

3.2.2 Effect of parameter b_0

To observe how the errors vary with b_0, parameter b_0 is changed from -2 to 2 while a_0 and p are set as $a_0=1$ and $p=0.05$.
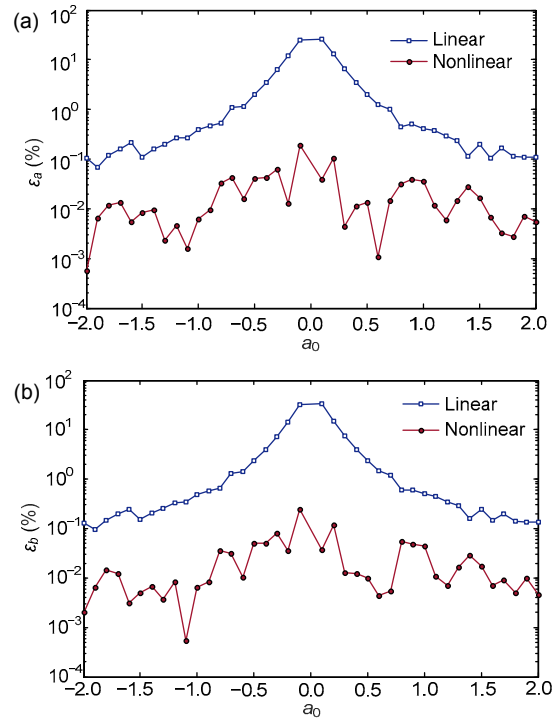

Fig. 1 Relative estimation errors between the estimated value a and its true value a_0 (a) and relative estimation errors between the estimated value b and its true value b_0 (b), with respect to parameter a_0
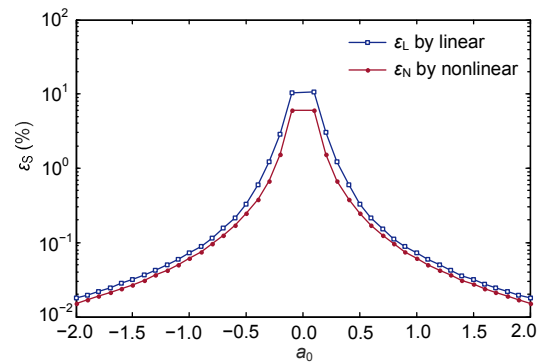

Fig. 2 Relative sum of squared errors with respect to parameter a_0

Figs. 3 and 4 show the relative estimation errors ε_a and ε_b, and the relative sum of squared errors ε_{L} and ε_{N}, respectively. Because parameter b_0 is in the exponential term, it has significant influence on the model. As b_0 increases, the magnitude of the signal in the measurements increases rapidly and the noise interference is reduced. As a consequence, the fitting error is also reduced.

Figs. 3 and 4 both demonstrate that the result of nonlinear fitting is superior to that of linear fitting.
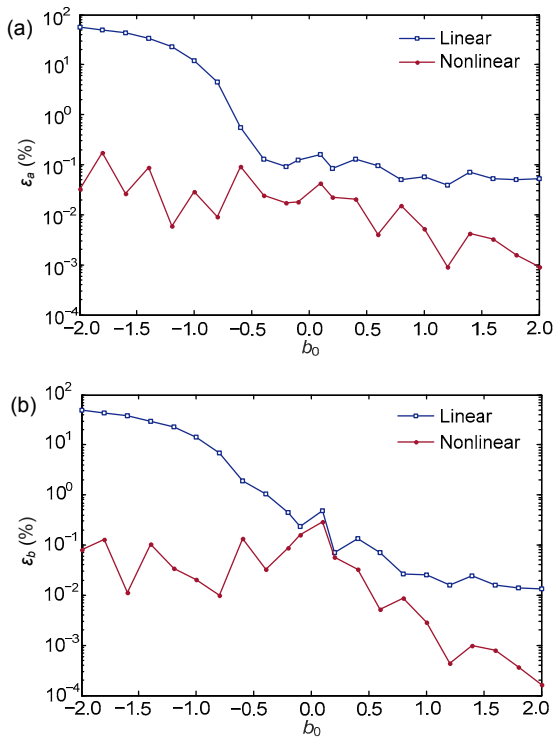
Fig. 3 Relative estimation errors between the estimated value a and its true value a_0 (a) and relative estimation errors between the estimated value b and its true value b_0 (b), with respect to parameter b_0
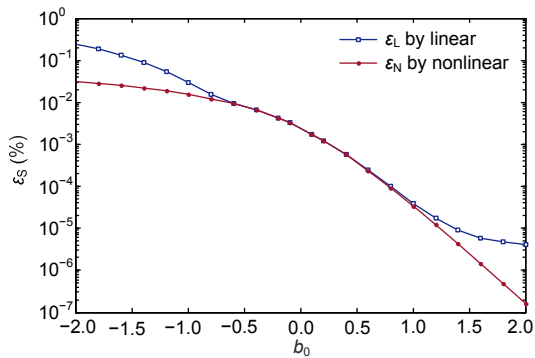

Fig. 4 Relative sum of squared errors with respect to parameter b_0

Because the difference between ε_L and ε_N is not obvious in Fig. 4 for $b_0 \in (-0.6, 0.6)$, the ratio between ε_L and ε_N (defined as $\varepsilon = \varepsilon_L/\varepsilon_N$) is plotted in Fig. 5, showing that ε_L is always larger than ε_N.

3.2.3 Effect of noise magnitude

To see the influence of noise magnitude on fitting, coefficient p is changed from 10^{-3} to 10^0, with $a_0 = 1$ and $b_0 = -1$.

As shown in Fig. 6, noise interference increases as p increases, and the relative estimation errors ε_a and ε_b in both methods tend to be higher. The deviation of the estimated parameter from the true value is higher for linear fitting than for nonlinear fitting. For example, when $p = 0.1$, the relative error between the estimated value a and the true value a_0 is about 23.93% for linear fitting, but only about 0.04% for nonlinear fitting. The relative error between the estimated value b and the true value b_0 is about 32.03% for linear fitting, but only about 0.02% for nonlinear fitting. Therefore, the nonlinear fitting method leads to more accurate estimates of unknown parameters.
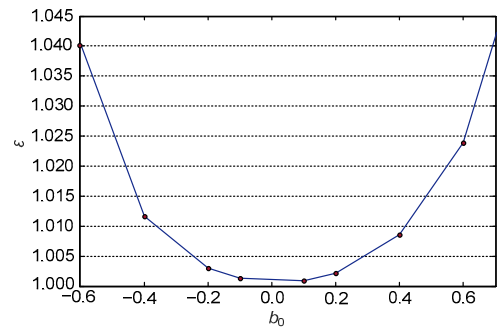

Fig. 5 Ratio between ε_L and ε_N ($\varepsilon = \varepsilon_L/\varepsilon_N$) with respect to b_0
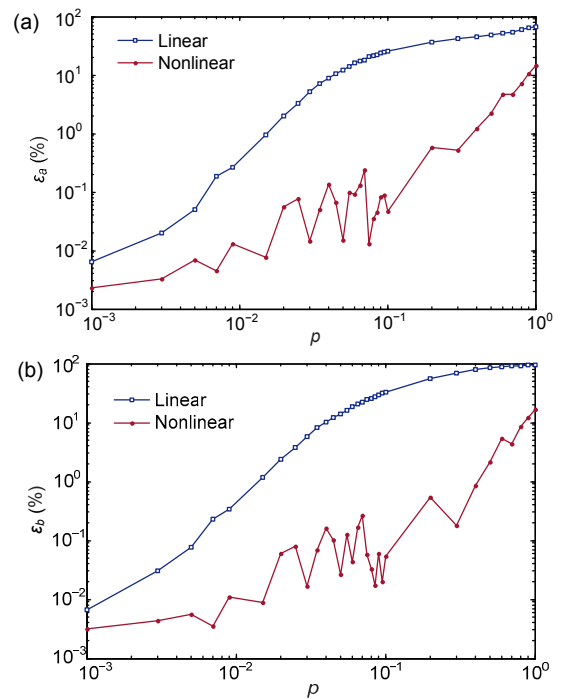

Fig. 6 Relative estimation errors between the estimated value a and its true value a_0 (a), and relative estimation errors between the estimated value b and its true value b_0 (b), with respect to parameter p

Fig. 7 shows the relative sum of squared errors ε_L and ε_N of two methods. Both tend to increase as p increases. Although Fig. 7 shows that the sums of squared errors of both methods are close to each other, the parameter estimation errors of two methods are notably different, as shown in Fig. 6.
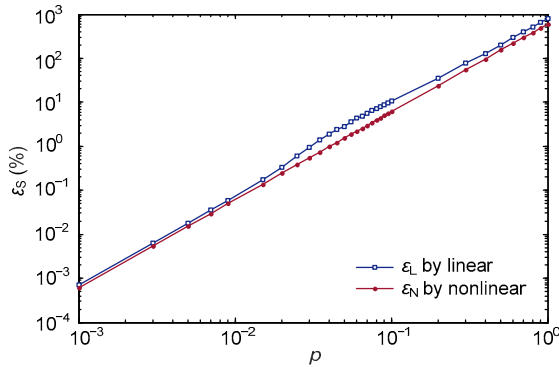

Fig. 7 Relative sum of squared errors with respect to parameter p

3.2.4 Effect of noise distribution

To see how the fitting is affected by noise distribution, the noise conforms to the uniform distribution in the interval of $[-\sqrt{3}, \sqrt{3}]$, with zero mean and variance of 1. Simulation is performed as in Section 3.2.3, when parameter p is changed.

As can be seen, Figs. 8 and 9 are quite similar to Figs. 6 and 7, respectively. The change of noise distribution does not have significant influence on the fitting accuracy.

4 Application

When light penetrates a water column, the spectral intensity of light is attenuated exponentially based on distance due to the absorption and scattering by water, following the Beer-Lambert law (Swinehart, 1962). The degree of light attenuation is characterized by the spectral attenuation coefficient of the water column, which is important for underwater image processing, ocean environment monitoring (Schechner and Karpel, 2004; Åhlén *et al.*, 2006), etc. Because the spectral attenuation coefficient differs among water columns, it is necessary to calibrate the spectral attenuation coefficient of water in situ.
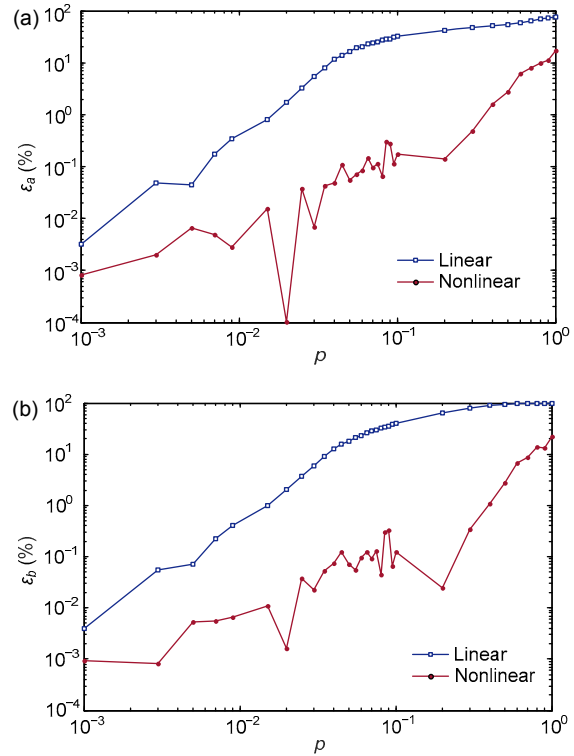

Fig. 8 Relative estimation errors between the estimated value a and its true value a_0 (a) and relative estimation errors between the estimated value b and its true value b_0 (b), with respect to parameter p when noise is uniformly distributed
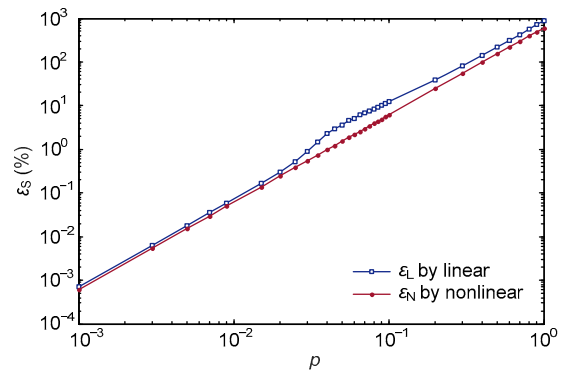

Fig. 9 Relative sum of squared errors with respect to parameter p when noise is uniformly distributed

Fig. 10 shows the schematic of an experimental setup for imaging underwater objects with narrowband color filters. The spectral attenuation coefficient of the water needs to be calibrated to recover the light power attenuated by water. There are commercial instruments dedicated to measurement of the spectral

attenuation coefficient of water. However, in our case, the coefficient is estimated directly based on underwater images to simplify the complexity of hardware.
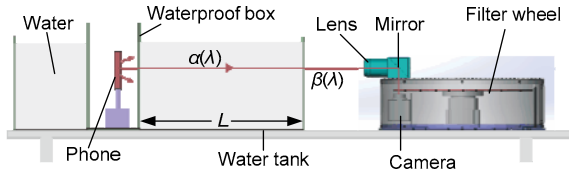

Fig. 10 Experimental setup for underwater imaging with narrowband color filters

The setup consists mainly of a mobile phone, water tank, and a spectral imaging system. The mobile phone (Xiaomi 3, Xiaomi, China) is placed in a waterproof glass box, with 60 color blocks displayed on its LCD screen (Fig. 11), and acts as the underwater object to be imaged. The waterproof box (35 cm×25 cm×50 cm) is made of 6-mm-thick quartz glass, placed in a water tank (300 cm×30 cm×30 cm) made of 10-mm-thick quartz glass. The water tank is filled with clean tap water. The spectral imaging system is placed outside the water tank, and consists mainly of an imaging lens, a set of narrowband color filters, and a monochrome charge-coupled device (CCD) camera. The filters (FB serious, Thorlabs, USA) are placed in a filter wheel, so that the filters can be changed by rotating the wheel. The central wavelength of the color filters is from 420 nm to 700 nm with an interval of 20 nm. The camera (Lm-165M, Lumenera, USA) has a resolution of 1392×1040 pixels and a bit depth of eight bits. The room temperature is about 20 °C.
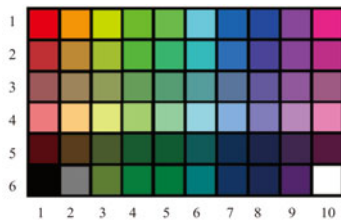

Fig. 11 Image on the mobile phone screen (References to color refer to the online version of this figure)

The spectral transmittance of the water can be expressed as (Guo *et al*., 2016)

$$k(\lambda,L) = \frac{I(\lambda,L)}{I_0(\lambda)} = \beta(\lambda)\mathrm{e}^{-\alpha(\lambda)L}, \quad (31)$$

where λ is the wavelength of light. $I_0(\lambda)$ and $I(\lambda, L)$ represent the brightness of a certain color batch in the image when the images are captured in air (i.e., without water in the light path) and captured underwater with an underwater distance of L, respectively. $\beta(\lambda)$ is the transmittance of the glass and $\alpha(\lambda)$ is the spectral attenuation coefficient of the water. The distance L, image brightness $I_0(\lambda)$ and $I(\lambda, L)$, and spectral transmittance $k(\lambda, L)$ can all be measured or calculated. $\beta(\lambda)$ and $\alpha(\lambda)$ are the unknown parameters to be estimated.

During the experiment, the underwater distance L is changed from 0.1 m to 2.6 m at an interval of 0.1 m by moving the waterproof box along the water tank. At each distance, images are captured with 15 color filters by rotating the filter wheel.

Table 1 shows the spectral transmittance of water with respect to L when the central wavelength of the color filters is 420 nm; i.e., only light in the wavelength range of 410–430 nm can pass the filter. Based on the exponential model in Eq. (31), the unknown parameters $\beta(\lambda)$ and $\alpha(\lambda)$ are estimated by the linear and nonlinear fitting methods as described in Section 2.

Table 1 Spectral transmittance of water

L (m)	k (a.u.)	L (m)	k (a.u.)
0.1	0.750	1.4	0.309
0.2	0.680	1.5	0.317
0.3	0.648	1.6	0.299
0.4	0.622	1.7	0.269
0.5	0.597	1.8	0.265
0.6	0.580	1.9	0.240
0.7	0.523	2.0	0.223
0.8	0.510	2.1	0.217
0.9	0.448	2.2	0.182
1.0	0.414	2.3	0.137
1.1	0.391	2.4	0.138
1.2	0.353	2.5	0.108
1.3	0.326	2.6	0.093

The parameter estimation results are summarized in Table 2 and shown in Fig. 12. It can be seen that the sum of squared errors using nonlinear fitting is only half of the linear value, showing high accuracy of the nonlinear method.

Table 2 Estimate of *α(λ)* and *β(λ)* obtained by both methods

Method	Estimate of $\beta(\lambda)$	Estimate of $\alpha(\lambda)$ (m^{-1})	Sum of squared errors
Linear	0.881	0.751	0.025
Nonlinear	0.815	0.673	0.013
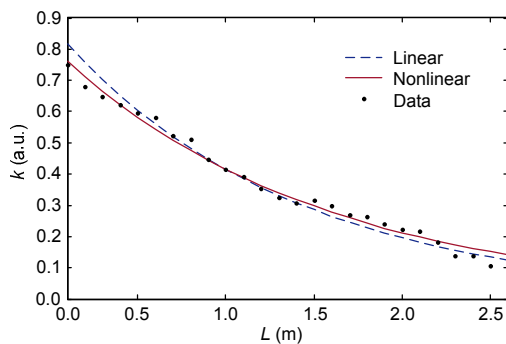
Fig. 12 Measurement data and fitting curves obtained by two methods

5 Conclusions

In this paper, the problem of parameter estimation in exponential models was discussed and two methods were considered, namely, linear and nonlinear fitting methods.

In the linear fitting method, the logarithmic value of the measurement data was fitted instead of the original data; in contrast, in the nonlinear method, the original data was fitted directly by the optimization algorithms. This discrepancy of the fitting data leads to a difference in the fitting goals and as a consequence, in the fitting results.

Theoretical analysis indicates that the estimates achieved by the linear method are simply the true values of a_0 and b_0 in an ideal case when the measurement data is not affected by noise. However, once noise exists in the measurement data, which is inevitable in practice, only suboptimal estimates can be obtained and the estimation error increases with the noise.

Simulation results expand the theoretical analysis and verify the superiority of nonlinear fitting in terms of fitting accuracy and robustness, especially for data with a low SNR.

Application of the fitting methods was also demonstrated in estimating the water spectral attenuation coefficient based on underwater images. Results show that the residual error by the nonlinear method is smaller than that by the linear method, indicating improvement in accuracy.

The results of this study can be helpful in parameter fitting when an exponential function is involved in the models.

References

Adi, N.S., 2015. Characterisation of the light environment and biophysical parameters of seagrass using remote sensing. PhD Thesis, The University of Queensland, Brisbane, Australia, p.23-38.
http://dx.doi.org/10.14264/uql.2015.998

Åhlén, J., Bengtsson, E., Sundgren, D., 2006. Evaluation of underwater spectral data for color correction applications. Proc. 5th WSEAS Int. Conf. on Circuits, Systems, Electronics, Control & Signal Processing, p.321-326.

Bardaji, R., Sánchez, A.M., Simon, C., *et al*., 2016. Estimating the underwater diffuse attenuation coefficient with a low-cost instrument: the KdUINO DIY buoy. *Sensors*, **16**(3):373-387. http://dx.doi.org/10.3390/s16030373

Floudas, C.A., Pardalos, P.M., 2001. Encyclopedia of Optimization. Kluwer Academic Publishers, Norwell, p.1129-1134. http://dx.doi.org/10.1007/0-306-48332-7

Guo, Y.L., Song, H., Liu, H.B., *et al*., 2016. Model-based restoration of underwater spectral images captured with narrowband filters. *Opt. Expr.*, **24**(12):13101-13120. http://dx.doi.org/10.1364/OE.24.013101

Hartley, H.O., 1961. The modified Gauss-Newton method for the fitting of nonlinear regression functions by least squares. *Technometrics*, **3**(2):269-280. http://dx.doi.org/10.2307/1266117

He, Q.R., Li, L.P., 2015. Algorithm study on reducing frequency measurement variance of acousto-optic spectrum analyzer. *Infrar. Laser Eng.*, **44**(5):1564-1568 (in Chinese).

Kaeli, J.W., Singh, H., Murphy, C., *et al*., 2011. Improving color correction for underwater image surveys. Proc. IEEE/MTS Oceans, p.805-810.

Lai, J.S., Yang, B., Lin, D.M., *et al*., 2013. The allometry of coarse root biomass: log-transformed linear regression or nonlinear regression. *PLoS One*, **8**(10):e77007. http://dx.doi.org/10.1371/journal.pone.0077007

Levenberg, K., 1944. A method for the solution of certain non-linear problems in least squares. *Q. Appl. Math.*, **2**:164-168. http://dx.doi.org/10.1090/qam/10666

Lieb, S.G., 1997. Simplex method of nonlinear least-squares: a logical complementary method to linear least-square analysis of data. *J. Chem. Educat.*, **74**(8):1008-1011. http://dx.doi.org/10.1021/ed074p1008

Liu, C.Q., Ma, J.S., Shao, X.L., *et al*., 2013. Detection of Gaussian beam distribution of light intensity. *Opt. Instrum.*, **35**(6):69-73 (in Chinese). http://dx.doi.org/10.3969/j.issn.1005-5630.2013.06.014

Marquardt, D.W., 1963. An algorithm for least-squares estimation of nonlinear parameters. *J. Soc. Ind. Appl. Math.*, **11**(2):431-441. http://dx.doi.org/10.1137/0111030

Mascaro, J., Litton, C.M., Hughes, R.F., *et al*., 2011. Minimizing bias in biomass allometry: model selection and log-transformation of data. *Biotropica*, **43**(6):649-653. http://dx.doi.org/10.1111/j.1744-7429.2011.00798.x

Novković, D., Nadderd, L., Kandić, A., *et al*., 2006. Testing the exponential decay law of gold [198]Au. *Nucl. Instrum. Methods Phys. Res. Sect. A*, **566**(2):477-480. http://dx.doi.org/10.1016/j.nima.2006.07.044

Peterson, P., Baker, E., McGaw, B., 2010. International Encyclopedia of Education. Elsevier Science, Oxford, p.339-346.

Schechner, Y.Y., Karpel, N., 2004. Clear underwater vision. Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, p.536-543. http://dx.doi.org/10.1109/CVPR.2004.1315078

Semkow, T.M., 2007. Exponential decay law and nuclear statistics. *In*: Semkow, T.M., Pommé, S., Jerome, S., *et al*. (Eds.), Applied Modeling and Computations in Nuclear Science. ACS, Washington DC. http://dx.doi.org/10.1021/bk-2007-0945.ch004

Simon, A., Shanmugam, P., 2016. Estimation of the spectral diffuse attenuation coefficient of downwelling irradiance in inland and coastal waters from hyperspectral remote sensing data: validation with experimental data. *Int. J. Appl. Earth Observ. Geoinform.*, **49**:117-125. http://dx.doi.org/10.1016/j.jag.2016.02.003

Swinehart, D.F., 1962. The Beer-Lambert law. *J. Chem. Educat.*, **39**(7):333-335. http://dx.doi.org/10.1021/ed039p333

Vickery, P.J., 2005. Simple empirical models for estimating the increase in the central pressure of tropical cyclones after landfall along the coastline of the United States. *J. Appl. Meteorol.*, **44**:1807-1826. http://dx.doi.org/10.1175/JAM2310.1

Wei, J.W., Lee, Z.P., 2013. Model of the attenuation coefficient of daily photosynthetically available radiation in the upper ocean. *Methods Oceanogr.*, **8**:56-74. http://dx.doi.org/10.1016/j.mio.2013.12.001

Xiao, X., White, E.P., Hooten, M.B., *et al*., 2011. On the use of log-transformation vs. nonlinear regression for analyzing biological power laws. *Ecology*, **92**(10):1887-1894. http://dx.doi.org/10.1890/11-0538.1