WILEY | Hindawi

Research Article

Efficient Parameters Estimation Method for the Separable Nonlinear Least Squares Problem

Ke Wang [ID],[1,2] **Guolin Liu** [ID],[1] **Qiuxiang Tao,**[1] **and Min Zhai**[1,2]

[1]*College of Geomatics, Shandong University of Science and Technology, Qingdao, Shandong 266590, China*
[2]*Schulich School of Engineering, University of Calgary, Calgary, Alberta T2N 1N4, Canada*

Correspondence should be addressed to Ke Wang; wke0115@126.com and Guolin Liu; gliu@sdust.edu.cn

In this work, we combine the special structure of the separable nonlinear least squares problem with a variable projection algorithm based on singular value decomposition to separate linear and nonlinear parameters. Then, we propose finding the nonlinear parameters using the Levenberg–Marquart (LM) algorithm and either solve the linear parameters using the least squares method directly or by using an iteration method that corrects the characteristic values based on the L-curve, according to whether or not the nonlinear function coefficient matrix is ill posed. To prove the feasibility of the proposed method, we compared its performance on three examples with that of the LM method without parameter separation. The results show that (1) the parameter separation method reduces the number of iterations and improves computational efficiency by reducing the parameter dimensions and (2) when the coefficient matrix of the linear parameters is well-posed, using the least squares method to solve the fitting problem provides the highest fitting accuracy. When the coefficient matrix is ill posed, the method of correcting characteristic values based on the L-curve provides the most accurate solution to the fitting problem.

1. Introduction

The separable nonlinear least squares problem is a special case of nonlinear least squares, and its model can be expressed as a linear combination of nonlinear functions. More generally, one can also consider that there are two sets of unknown parameters, where one set is dependent on the other and can be explicitly eliminated. This problem was first proposed in parameter estimation of the atomic physics particle half-life formula developed by Golub and Pereyra in 1973 [1]. In real life, models of this type are very common. For example, in the machine learning community, neural networks, their numerous variants [2, 3], and some neuro fuzzy systems [4, 5] take the form of a linear ensemble of some nonlinear basis functions. In the field of signal processing, Prony's method [6, 7], which takes the sum of complex exponentials, is frequently used to analyze the frequency components of a signal. In the field of algorithm application, waveform data decomposition is one of the key steps in processing based on airborne full-waveform light

detection and ranging (LiDAR) data. The full waveform can be decomposed into a linear combination of multiple Gaussian functions. Through waveform data decomposition, discrete point cloud and waveform parameter information can be obtained [8]. Furthermore, this approach has many applications in areas such as mechanical systems [9], telecommunications [10], robotics [11], and environmental sciences [12]. These applications can be viewed as nonlinear data-fitting problems in terms of numerical expression. However, data-fitting problems are often quite challenging numerically. Fortunately, by exploiting the special structure of separable nonlinear models, efficient algorithms can be obtained.

Based on the special structure of the separable nonlinear least squares problem, Golub and Pereyra [1] proposed the variable projection (VP) algorithm to eliminate linear parameters and obtain simplified functions involving only nonlinear parameters and also used the Levenberg–Marquart (LM) algorithm for its solution. Dimensionality reduction of parameter space improves the possibility of

obtaining a global optimal solution [13]. Therefore, improvement and application [14] of VP have been conducted on that basis. Kaufman [15] proposed a modified VP algorithm based on trapezoidal decomposition to simplify the calculation, which reduced its computational complexity and improved computational efficiency. Subsequently, Ruano et al. [16] proposed an improved VP algorithm based on QR decomposition for the sparse case of nonlinear function matrix, which also effectively improved operation efficiency. Further, Gan and Li [17] proposed a VP algorithm based on Gram–Schmidt matrix decomposition for the case in which the number of observations is much larger than the number of linear parameters, which reduces the amount of calculation required.

In this study, in view of the ill-posed condition of a nonlinear function matrix, singular value decomposition (SVD) [18] is adopted to simplify the VP algorithm and improve the stability of the calculation. Then, after linear parameters are eliminated by the improved VP algorithm, the separable least squares problem is transformed into an optimization problem containing only nonlinear parameters [19]. As regards an estimation method for nonlinear parameters, Liu et al. [20] combined this with sequential quadratic programming (SQP), developing a gradient-based optimization algorithm to determine the optimal time-delays and system parameters in a novel dynamic optimization problem for nonlinear multistage systems with time delays. However, this approach was restricted to parameter identification problems.

The common nonlinear least squares iterative solutions [21] are the gradient descent method [22], Gauss–Newton method [23], and Levenberg–Marquart (LM) method [24, 25]. For example, in [26], the nonlinear least squares problem of the distributional robust parameter identification model for time-delay systems is transformed into a single-level optimization problem and a gradient-based optimization method is developed to solve the transformed problem. The method only involves the first-order moment information and is simple to calculate. To obtain robust estimates against the noise in measurements, Liu et al. [27] proposed a robust estimation formulation, in which the cost function was the variance of the error function and an additional constraint indicates an allowable sacrifice from the optimal expectation value of the classical estimation problem. On this basis, a gradient-based optimization algorithm to numerically solve the classical and robust parameter estimation problems was developed. This is more efficient than the existing methods used for problems, where optimization parameters outnumber constraints. This method involves simple calculations, but its convergence speed is generally slower than that of the Gauss–Newton method.

Torres et al. [28] used the sequentially semiseparable matrix to calculate the Jacobian matrix [29] and Hessian matrix [30], employing the Gauss–Newton method to optimize the output error of the global system. The effectiveness of the algorithm was verified by numerical examples. Bellavia et al. [31] improved the approximation function by controlling the accuracy level when the accuracy was too low to be optimized, and then proposed the LM method based on the dynamic precision relationship between the evaluation function and gradient for solving large-scale nonlinear least squares problems. They proved the global and local convergence and complexity of this method. The Gauss–Newton method has the advantages of fast convergence and high precision. However, the Jacobian matrix is required to be of full rank in the iterative process. If the problem has high nonlinearity or the residual is large, the method may not produce convergence. Therefore, the LM algorithm overcomes this shortcoming and adjusts the damping parameters according to the idea of a trust region, to effectively control the direction of iterative descent. Once the nonlinear parameters were determined, the least squares (LS) method [32] is used to estimate the linear parameters.

In view of a general nonlinear multistage system with time-delay and system parameters, Liu et al. [33] proposed a new parameter estimation formulation, in which the cost function is the variance of the error function and the constraint indicates an allowable sacrifice from the optimal expectation value of the classical parameter estimation problem. This parameter estimation approach is capable of solving parameter estimation problems with multiple stages and multiple time delays and, compared with classical parameter estimation, it is able to withstand the uncertainty in the distribution of measurement data. Nevertheless, this method has the limitation of relying on the statistical distribution of the noisy measurement output.

To enhance the estimation accuracy, Ding et al. [34, 35] presented a filtering-based gradient iterative algorithm and a filtering-based least squares iterative algorithm, which resulted in improved convergence speed. However, when nonlinear parameter estimation led to the ill condition of the linear least squares coefficient matrix, the estimation results of the linear parameters obtained by the LS method were unstable and even sometimes significantly different from the true values. For this problem, the regularization method is often used to solve for ill-posed problems in linear parameter estimation [36], such as the Tikhonov regularization method [37], truncated singular value method [38], and iteration by correcting characteristic values [39, 40].

In addition, Chen et al. [41] proposed a weighted generalized crossvalidation method to determine Tikhonov regularization parameters for the regularization of separable nonlinear least squares ill-posed problems based on the VP algorithm and verified its effectiveness experimentally. Aiming at the randomness of parameter selection in the iteration by correcting characteristic values in the process of linear least squares solution, Zhai et al. [42] constructed the L-curve [43, 44] of the relationship between the norm of the parameter solution and the residual. They selected the maximum curvature point as the regularization parameter and verified the correctness of the method through numerical experiments.

The existing literature presents several effective solutions for the parameter estimation problem, but only a few studies have been conducted on the structural transformation of separable nonlinear models. In this study, we

consider the special structure of the separable nonlinear least squares problem, separating two types of parameters using a VP algorithm based on SVD. We then use the LM algorithm to estimate the nonlinear parameters while using the LS method or iteration by correcting characteristic values based on the L-curve method. By comparing the experimental results of parameter estimation of the Gauss function fitting model and fractional fitting model with those of the LM method without parameter separation, the validity of the VP algorithm based on SVD is evaluated and the accuracy of different linear parameter estimation methods is analyzed.

The remainder of this paper is outlined as follows. In Section 2, based on an improved VP algorithm derived from SVD, the methods of nonlinear parameter estimation and linear parameter estimation are explained and the algorithm for solving separable nonlinear least squares problems is provided. In Section 3, the Gaussian function fitting model and the fractional fitting model experiments are used to compare and analyze the proposed method with the traditional LM algorithm without unseparated parameters. Finally, we present our conclusions in Section 4.

2. Solution of Parameters for Separable Nonlinear Least Squares

2.1. Variable Projection Algorithm Based on SVD. Consider a set of observations (t_i, y_i) $(i = 1, 2, \ldots, m)$. The problem of parameter estimation is to find the optimal parameters $\widehat{a} = (a_1, a_2, \ldots, a_p)^{\mathrm{T}}$ and $\widehat{b} = (b_1, b_2, \ldots, b_q)^{\mathrm{T}}$ when the following formula reaches the minimum value:

$$(\widehat{a}, \widehat{b}) = \arg\min_{a,b} \sum_{i=1}^{m} \left[y(t_i) - \sum_{j=1}^{p} a_j \phi_j (\mathbf{b}; t_i) \right]^2, \quad (1)$$

where $\varphi_j (b; t_i)$ $(i = 1, 2, \ldots, m; j = 1, 2, \ldots, p)$ is a nonlinear function and p and q are the number of linear and nonlinear parameters to be estimated, respectively. The above formula can be written in a matrix form as

$$(\widehat{a}, \widehat{b}) = \arg\min_{a,b} \sum_{i=1}^{m} [\mathbf{y} - \mathbf{\Phi}(\mathbf{b})\mathbf{a}]^2 = \arg\min_{a,b} \|\mathbf{y} - \mathbf{\Phi}(\mathbf{b})\mathbf{a}\|_2^2, \quad (2)$$

where the column of $\mathbf{\Phi}(\mathbf{b})$ corresponds to the nonlinear function $\varphi_j (\mathbf{b}; \mathbf{t})$ associated with the parameter $\mathbf{b}$ and $\|\cdot\|_2$ denotes the Euclidean norm.

Let

$$f(\mathbf{a}, \mathbf{b}) = \|\mathbf{y} - \mathbf{\Phi}(\mathbf{b})\mathbf{a}\|_2^2. \quad (3)$$

For the given nonlinear parameters $\mathbf{b}$, the linear parameters $\widehat{a}$ can be estimated by solving the following linear least squares problem:

$$\widehat{a} = \left(\mathbf{\Phi}^{\mathrm{T}}(\mathbf{b})\mathbf{\Phi}(\mathbf{b}) \right)^{-1} \mathbf{\Phi}(\mathbf{b})\mathbf{y} = \mathbf{\Phi}^{+}(\mathbf{b})\mathbf{y}, \quad (4)$$

where $\mathbf{\Phi}^{+}(\mathbf{b})$ is the pseudoinverse of $\mathbf{\Phi}(\mathbf{b})$. Inserting (4) into (3),

$$\widehat{b} = \arg\min_{a,b} \|\mathbf{y} - \mathbf{\Phi}(\mathbf{b})\mathbf{\Phi}^{+}(\mathbf{b})\mathbf{y}\|_2^2$$

$$= \arg\min_{a,b} \|(\mathbf{I} - \mathbf{\Phi}(\mathbf{b})\mathbf{\Phi}^{+}(\mathbf{b}))\mathbf{y}\|_2^2$$

$$= \arg\min_{a,b} \|(\mathbf{I} - \mathbf{P}_{\mathbf{\Phi}(\mathbf{b})})\mathbf{y}\|_2^2$$

$$= \arg\min_{a,b} \|\mathbf{P}_{\mathbf{\Phi}(\mathbf{b})}^{\perp}\mathbf{y}\|_2^2, \quad (5)$$

where $\mathbf{P}_{\mathbf{\Phi}(\mathbf{b})}$ is the orthogonal projector on the linear space spanned by the columns of $\mathbf{\Phi}(\mathbf{b})$ and $\mathbf{P}_{\mathbf{\Phi}(\mathbf{b})}^{\perp}$ is the projector on the orthogonal complement of the column space.

To simplify the calculation, the matrix $\mathbf{\Phi}(\mathbf{b})$, which is composed of nonlinear functions, is decomposed by SVD:

$$\mathbf{\Phi}(\mathbf{b}) = \mathbf{U}\mathbf{S}\mathbf{V}^{\mathrm{T}}, \quad (6)$$

where $\mathbf{U}$ is an $m \times m$ orthogonal matrix, $\mathbf{S}$ is an $m \times p$ diagonal matrix, and $\mathbf{V}$ is a $p \times p$ orthogonal matrix. We obtain $\mathbf{\Phi}^{+}(\mathbf{b})$ as

$$\mathbf{\Phi}^{+}(\mathbf{b}) = \mathbf{V}\mathbf{S}^{-1}\mathbf{U}^{\mathrm{T}}. \quad (7)$$

Then,

$$\mathbf{P}_{\mathbf{\Phi}(\mathbf{b})} = \mathbf{\Phi}(\mathbf{b})\mathbf{\Phi}^{+}(\mathbf{b}) = \mathbf{U}\mathbf{S}\mathbf{V}^{\mathrm{T}}\mathbf{V}\mathbf{S}^{-1}\mathbf{U}^{\mathrm{T}} = \mathbf{U}\begin{bmatrix} \mathbf{I}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\mathbf{U}^{\mathrm{T}}. \quad (8)$$

Suppose the rank of the matrix $\mathbf{\Phi}(\mathbf{b})$ is r, the first r elements on the diagonal of $\mathbf{S}$ are not zero, i.e., $r \leq p$. $\mathbf{U}$ can be divided into $m \times r$ matrix $\mathbf{U}_1$ and $m \times (m - r)$ matrix $\mathbf{U}_2$. $\mathbf{V}$ can be divided into $p \times r$ matrix $\mathbf{V}_1$ and $p \times (p - r)$ matrix $\mathbf{V}_2$. Then,

$$\mathbf{\Phi}(\mathbf{b}) = \begin{bmatrix} \underbrace{\mathbf{U}_1}_{m \times r} & \underbrace{\mathbf{U}_2}_{m \times (m-r)} \end{bmatrix} \begin{bmatrix} \mathbf{S}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \underbrace{\mathbf{V}_1^{\mathrm{T}}}_{r \times p} \\ \underbrace{\mathbf{V}_2^{\mathrm{T}}}_{(p-r) \times p} \end{bmatrix},$$

$$\mathbf{P}_{\mathbf{\Phi}(\mathbf{b})}^{\perp} = \mathbf{I} - \begin{bmatrix} \underbrace{\mathbf{U}_1}_{m \times r} & \underbrace{\mathbf{U}_2}_{m \times (m-r)} \end{bmatrix} \begin{bmatrix} \mathbf{I}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \underbrace{\mathbf{U}_1^{\mathrm{T}}}_{r \times m} \\ \underbrace{\mathbf{U}_2^{\mathrm{T}}}_{(m-r) \times m} \end{bmatrix} \quad (9)$$

$$= -\begin{bmatrix} \underbrace{\mathbf{U}_1}_{m \times r} & \underbrace{\mathbf{U}_2}_{m \times (m-r)} \end{bmatrix} \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{p-r} \end{bmatrix} \begin{bmatrix} \underbrace{\mathbf{U}_1^{\mathrm{T}}}_{r \times m} \\ \underbrace{\mathbf{U}_2^{\mathrm{T}}}_{(m-r) \times m} \end{bmatrix}$$

$$= \mathbf{U}_2 \mathbf{U}_2^{\mathrm{T}}.$$

The matrix composed of m residual functions is simplified to the following equation by VP based on SVD:

$$\mathbf{F}(\mathbf{b}) = \mathbf{P}_{\mathbf{\Phi}(\mathbf{b})}^{\perp}\mathbf{y} = \mathbf{U}_2 \mathbf{U}_2^{\mathrm{T}}\mathbf{y}. \quad (10)$$

Then, the objective function of the separable nonlinear least squares problem is simplified to

$$f(\mathbf{b}) = \|\mathbf{F}(\mathbf{b})\|_2^2 = \left\|\mathbf{P}_{\Phi(\mathbf{b})}^{\perp}\mathbf{y}\right\|_2^2 = \left\|\mathbf{U}_2\mathbf{U}_2^{\mathrm{T}}\mathbf{y}\right\|_2^2. \tag{11}$$

Theorem 1. *We assume that in the open set $\Omega \in \mathfrak{R}^k$, $\Phi(\mathbf{b})$ has constant rank $r = \min(p, q)$.*

(a) *If $\widehat{b}$ is a critical point (or a global minimizer in Ω) of $f(\mathbf{b})$, and combined with (4), then $(\widehat{a}, \widehat{b})$ is a critical point of $f(\mathbf{a}, \mathbf{b})$ (or a global minimizer for $\mathbf{b} \in \Omega$) and $f(\mathbf{a}, \mathbf{b}) = f(\mathbf{b})$*

(b) *If $(\widehat{a}, \widehat{b})$ is a global minimizer of $f(\mathbf{a}, \mathbf{b})$ for $\mathbf{b} \in \Omega$, then $\widehat{b}$ is a global minimizer of $f(\mathbf{b})$ in Ω and $f(\mathbf{b}) = f(\mathbf{a}, \mathbf{b})$. Furthermore, if there is a unique $\widehat{b}$ among the minimizing pairs of $f(\mathbf{a}, \mathbf{b})$, then $\widehat{a}$ must satisfy (4).*

Proof of Theorem 1. From (3) we see that $f(\mathbf{a}, \mathbf{b}) = \|\mathbf{y} - \Phi(\mathbf{b})\mathbf{a}\|_2^2$. Therefore,

$$\frac{1}{2}\nabla f(\widehat{a}, \widehat{b}) = -\left(\mathbf{P}_{\Phi(\mathbf{b})}^{\perp}\mathbf{y}\right)^{\mathrm{T}}\left(\Phi(\mathbf{b})\Phi^{+}(\mathbf{b})\mathbf{y}\oplus\Phi(\mathbf{b})\right), \tag{12}$$

where $\oplus$ signifies a direct sum.

Assume that $\widehat{b}$ is a global minimizer of $f(\mathbf{b})$ in Ω and $\mathbf{a}$ is defined by (4); then,

$$\frac{1}{2}\nabla f(\widehat{a}, \widehat{b}) = -\left(\mathbf{P}_{\Phi(\mathbf{b})}^{\perp}\mathbf{y}\right)^{\mathrm{T}}\left(\Phi(\mathbf{b})\Phi^{+}(\mathbf{b})\mathbf{y}\oplus\Phi(\mathbf{b})\right) = \frac{1}{2}\nabla f(\widehat{b})\oplus\mathbf{0}. \tag{13}$$

$$\mathbf{b}_{k+1} = \mathbf{b}_k + \alpha_k\mathbf{d}_k, \tag{14}$$

Because $\mathbf{y}^{\mathrm{T}}\mathbf{P}_{\Phi(\mathbf{b})}^{\perp}\Phi(\mathbf{b}) = \mathbf{0}$, then $(\widehat{a}, \widehat{b})$ is a critical point of $f(\mathbf{a}, \mathbf{b})$.

Assume that $\widehat{b}$ is a global minimizer of $f(\mathbf{b})$ in Ω and $\widehat{a}$ satisfies (4). Then, clearly, $f(\widehat{a}, \widehat{b}) = f(\widehat{b})$. If we assume that there exists $(\mathbf{a}^{*}, \mathbf{b}^{*}), \mathbf{a}^{*} \in \Omega$, such that $f(\mathbf{a}^{*}, \mathbf{b}^{*}) < f(\widehat{a}, \widehat{b})$. Because for any $\mathbf{b}$ we have $f(\mathbf{b}) \leq f(\mathbf{a}, \mathbf{b})$, then it follows that $f(\mathbf{b}^{*}) \leq f(\mathbf{a}^{*}, \mathbf{b}^{*}) < f(\widehat{a}, \widehat{b}) = f(\widehat{b})$, which is a contradiction to the fact that $\widehat{b}$ is a global minimizer of $f(\mathbf{b})$ in Ω. Therefore, $(\widehat{a}, \widehat{b})$ is a global minimizer of $f(\mathbf{a}, \mathbf{b})$ in Ω, and part (a) of the theorem is proved.

Conversely, suppose that $(\widehat{a}, \widehat{b})$ is a global minimizer of $f(\mathbf{a}, \mathbf{b})$ in Ω. Then, as mentioned above $f(\widehat{b}) \leq f(\widehat{a}, \widehat{b})$. Now, let $\mathbf{a}^{*} = \Phi^{+}(\widehat{b})\mathbf{y}$. Then, we have $f(\widehat{b}) = f(\mathbf{a}^{*}, \widehat{b}) \leq f(\widehat{a}, \widehat{b})$; however, because $(\widehat{a}, \widehat{b})$ is a global minimizer, we must have equality. If there is a unique $\mathbf{a}$ among the minimizers of $f(\mathbf{a}, \mathbf{b})$, then $\mathbf{a}^{*} \equiv \widehat{a}$. We still have that $\widehat{b}$ is a global minimizer of $f(\widehat{b})$. Assume that it is not. Thus, there will be $\overline{b} \in \Omega$, such that $f(\overline{b}) < f(\widehat{b})$. Let $\overline{a}$ be equal to $\Phi^{+}(\overline{b})\mathbf{y}$. Then, $f(\overline{b}) = f(\overline{a}, \overline{b}) < f(\widehat{b}) = f(\widehat{a}, \widehat{b})$, which is a contradiction to the fact that $(\widehat{a}, \widehat{b})$ is a global minimizer of $f(\mathbf{a}, \mathbf{b})$. This completes the proof. □

where $\mathbf{b}_k$ is current nonlinear parameter vector. α_k is a small step length that ensures the decrease of the objection function (11), which is calculated by an imprecise search method, such as line search, in which we let m_k be the smallest nonnegative integer m satisfying $f(\mathbf{x}_k + \rho^m\mathbf{d}_k) \leq f(\mathbf{x}_k) + \sigma\rho^m\mathbf{g}_k^{\mathrm{T}}\mathbf{d}_k$ in the kth iteration process. Then,

$$\alpha_k = \rho^{m_k}, \tag{15}$$

where $\mathbf{d}_k$ is search direction, which can be determined by the following equation:

$$\mathbf{d}_k = \left(\mathbf{J}_k^{\mathrm{T}}\mathbf{J}_k + \mu_k\mathbf{I}\right)^{-1}\mathbf{J}_k\mathbf{F}_k, \tag{16}$$

where $\mathbf{J}_k \in \mathfrak{R}^{m\times p}$ is the Jacobian matrix of $\mathbf{F}(\mathbf{b})$. Kaufman [15] formulated an explicit analytic expression for the Jacobian. This ensures the efficiency and reliability of the VP algorithms. In addition, combined with the simplified objective function by SVD, the Jacobian matrix $\mathbf{J}_k$ of the residual vector $\mathbf{F}$ is expressed as

$$\begin{aligned}\mathbf{J}_k = \mathbf{F}(\mathbf{b}) &= \left(\mathbf{P}_{\Phi(\mathbf{b})}^{\perp}\mathbf{y}\right) \\ &= -\mathbf{P}_{\Phi(\mathbf{b})}^{\perp}(\Phi(\mathbf{b}))\Phi^{+}(\mathbf{b})\mathbf{y} \\ &= -\mathbf{U}_2\mathbf{U}_2^{\mathrm{T}}(\Phi(\mathbf{b}))\mathbf{V}_1\mathbf{S}_r^{-1}\mathbf{U}_1^{\mathrm{T}}\mathbf{y},\end{aligned} \tag{17}$$

2.2. Nonlinear Parameter Estimation Using the LM Algorithm. For a separable least squares problem with only nonlinear parameters, the LM algorithm is adopted for the solution and the nonlinear parameters $\mathbf{b}$ is updated as

where μ_k is damping parameter. It is adjusted with a strategy similar to the trust region radius [45], and a quadratic function is defined at the current iteration point as follows:

$$v(\mathbf{d}) = f(\mathbf{x}_k) + \left(\mathbf{J}_k^{\mathrm{T}}\mathbf{F}_k\right)^{\mathrm{T}}\mathbf{d} + \frac{1}{2}\mathbf{d}^{\mathrm{T}}\left(\mathbf{J}_k^{\mathrm{T}}\mathbf{J}_k\right)^{\mathrm{T}}\mathbf{d}. \tag{18}$$

Then, the incremental ratio of $v(\mathbf{x}_k + \mathbf{d}_k)$ to the objective function is calculated:

$$\eta_k = \frac{\Delta f(\mathbf{d}_k)}{\Delta v(\mathbf{d}_k)} = \frac{f(\mathbf{x}_{k+1}) - f(\mathbf{x}_k)}{\left(\mathbf{J}_k^{\mathrm{T}}\mathbf{F}_k\right)^{\mathrm{T}}\mathbf{d}_k + (1/2)\mathbf{d}_k^{\mathrm{T}}\left(\mathbf{J}_k^{\mathrm{T}}\mathbf{J}_k\right)^{\mathrm{T}}\mathbf{d}_k}. \tag{19}$$

When η_k is close to one, the fitting between the quadratic function $v(\mathbf{d})$ and objective function is good at point $\mathbf{x}_k$. LM is used to solve the nonlinear least squares problem. The parameter μ should be smaller; that is, in this case, the Gauss–Newton method is more effective. When it is close to zero, the fitting between the quadratic function $v(\mathbf{d})$ and objective function is poor at point $\mathbf{x}_k$. LM is used to solve the nonlinear least squares problem, and the parameter μ should be larger. When η_k is neither close to zero nor one, then μ_k is suitable and does not need to be adjusted. The critical values of η are usually 0.25 and 0.75, and the adjustment rules of μ are as follows.

When η_k is close to one, the fitting between quadratic function $v(\mathbf{d})$ and the objective function is good at point $\mathbf{x}_k$. The parameter μ should be smaller when the LM algorithm is used to solve the nonlinear least squares problem. When η_k is close to zero, the fitting between the quadratic function $v(\mathbf{d})$ and the objective function is poor at point $\mathbf{x}_k$. The parameter μ should be larger when LM is used to solve the nonlinear least squares problem. When η_k is neither close to zero nor one, μ is suitable and does not need to be adjusted. The critical values are usually 0.25 and 0.75, and the adjustment rules of μ are as follows:

$$\mu_{k+1} = \begin{cases} 0.1\mu_k, & \eta_k > 0.75, \\ \mu_k, & 0.25 \le \eta_k \le 0.75, \\ 10\mu_k, & \eta_k < 0.25. \end{cases} \tag{20}$$

2.3. Linear Parameter Estimation by Correcting Characteristic Value Based on L-Curve. The nonlinear parameter estimations $\widehat{b}$ are solved in Section (2); the linear parameters $\widehat{a}$ can then be calculated by (4). When $\Phi^{\mathrm{T}}(\mathbf{b})\Phi(\mathbf{b})$ is a nonsingular matrix, the least squares method is used to calculate it directly. However, when $\Phi^{\mathrm{T}}(\mathbf{b})\Phi(\mathbf{b})$ is singular or the condition number of $\Phi(\mathbf{b})$ exceeds 100, solution of the linear least squares method is unstable or cannot be solved. Therefore, it needs to be solved by the regularization method. In this study, an iteration method that corrects characteristic values based on the L-curve is used to solve the problem.

The least squares normal equation for the linear parameter estimation is

$$\Phi^{\mathrm{T}}(\mathbf{b})\Phi(\mathbf{b})\widehat{a} = \Phi^{\mathrm{T}}(\mathbf{b})\mathbf{y}. \tag{21}$$

Adding $\lambda\mathbf{a}$ to both sides of (21), we obtain

$$\left(\Phi^{\mathrm{T}}(\mathbf{b})\Phi(\mathbf{b}) + \lambda\mathbf{I}\right)\widehat{a} = \Phi^{\mathrm{T}}(\mathbf{b})\mathbf{y} + \lambda\widehat{a}. \tag{22}$$

Let $\mathbf{N} = \Phi^{\mathrm{T}}(\mathbf{b})\Phi(\mathbf{b})$ and $\mathbf{W} = \Phi(\mathbf{b})\mathbf{y}$, the step of the iteration method that corrects the characteristic value is

$$\begin{aligned} \widehat{a}_{k+1} &= (\mathbf{N} + \lambda\mathbf{I})^{-1}\left(\mathbf{W} + \lambda\widehat{a}_k\right) \\ &= \left((\mathbf{N} + \lambda\mathbf{I})^{-1} + \left((\mathbf{N} + \lambda\mathbf{I})^{-1}\right)^2 + \cdots + \left((\mathbf{N} + \lambda\mathbf{I})^{-1}\right)^{k+1}\right) \\ &\quad \mathbf{W} + \left((\mathbf{N} + \lambda\mathbf{I})^{-1}\right)^{k+1}\widehat{a}_k, \end{aligned} \tag{23}$$

where $\widehat{a}_k$ is the current nonlinear parameter vector and λ is a regularization parameter selected according to the L-curve method. The L-curve describes the relationship between the norm of the regularized solution $\|\mathbf{a}(\lambda)\|_2^2$ and the residual norm $\|\mathbf{y} - \Phi(\mathbf{b})\widehat{a}(\lambda)\|_2^2$, which corresponds to each set of regularized parameter values.

The convergence of the spectral correction iterative method based on the L-curve is proved as follows.

Let the initial estimated value of $\mathbf{a}$ be $\widehat{a}_0$, $\mathbf{N}_{\mathrm{LP}} = (\mathbf{N} + \lambda\mathbf{I})^{-1}$. The iterative calculation process of spectral correction can then be written as

$$\begin{aligned} \widehat{a}_1 &= (\mathbf{N} + \lambda\mathbf{I})^{-1}\mathbf{W} + (\mathbf{N} + \lambda\mathbf{I})^{-1}\widehat{a}_0 = \mathbf{N}_{\mathrm{LP}}\mathbf{W} + \mathbf{N}_{\mathrm{LP}}\widehat{a}_0, \\ \widehat{a}_2 &= \mathbf{N}_{\mathrm{LP}}\mathbf{W} + \mathbf{N}_{\mathrm{LP}}\widehat{a}_1 = \mathbf{N}_{\mathrm{LP}}\mathbf{W} + \mathbf{N}_{\mathrm{LP}} \cdot \mathbf{N}_{\mathrm{LP}}\mathbf{W} + \mathbf{N}_{\mathrm{LP}} \cdot \mathbf{N}_{\mathrm{LP}}\widehat{a}_0, \\ &\vdots \\ \widehat{a}_i &= \mathbf{N}_{\mathrm{LP}}\mathbf{W} + \mathbf{N}_{\mathrm{LP}}\widehat{a}_{i-1} = \mathbf{N}_{\mathrm{LP}}\mathbf{W} + \mathbf{N}_{\mathrm{LP}} \cdot \mathbf{N}_{\mathrm{LP}}\mathbf{W} \\ &\quad + \cdots + \underbrace{\mathbf{N}_{\mathrm{LP}} \cdots \mathbf{N}_{\mathrm{LP}}}_{i}\mathbf{W} + \underbrace{\mathbf{N}_{\mathrm{LP}} \cdots \mathbf{N}_{\mathrm{LP}}}_{i}\widehat{a}_0. \end{aligned} \tag{24}$$

When $i = m$, m is the number of inadequate iterations of the spectral modified iteration method; therefore, the relationship between estimated value $\widehat{a}_m$ and iteration initial value $\widehat{a}_0$ is

$$\begin{aligned} \widehat{a}_m &= \left(\mathbf{N}_{\mathrm{LP}} + \mathbf{N}_{\mathrm{LP}} \cdot \mathbf{N}_{\mathrm{LP}} + \cdots + \underbrace{\mathbf{N}_{\mathrm{LP}} \cdots \mathbf{N}_{\mathrm{LP}}}_{m}\right)\mathbf{W} \\ &\quad + \underbrace{\mathbf{N}_{\mathrm{LP}} \cdots \mathbf{N}_{\mathrm{LP}}}_{m}\widehat{a}_0. \end{aligned} \tag{25}$$

When $i = j$, j is the number of sufficient iterations of the spectral modified iterative method, and $m < j$; therefore, the relationship between estimated value $\widehat{a}_m$ and full iteration result $\widehat{a}_0$ is

$$\begin{aligned} \widehat{a}_m &= \underbrace{\mathbf{N}_{\mathrm{LP}} \cdots \mathbf{N}_{\mathrm{LP}}}_{m-j}\widehat{a}_j - \underbrace{\mathbf{N}_{\mathrm{LP}} \cdots \mathbf{N}_{\mathrm{LP}}}_{m-j} \\ &\quad \left(\mathbf{N}_{\mathrm{LP}} + \mathbf{N}_{\mathrm{LP}} \cdots \mathbf{N}_{\mathrm{LP}} + \cdots + \underbrace{\mathbf{N}_{\mathrm{LP}} \cdots \mathbf{N}_{\mathrm{LP}}}_{j-m}\right)\mathbf{W}. \end{aligned} \tag{26}$$

According to (23), the relation between the estimated result $\widehat{a}_j$ of spectral correction and the initial iteration value $\widehat{a}_0$ is as follows:

$$\hat{a}_i = \left(\mathbf{N}_{\text{LP}} + \mathbf{N}_{\text{LP}} \cdot \mathbf{N}_{\text{LP}} + \cdots + \underbrace{\mathbf{N}_{\text{LP}} \cdots \mathbf{N}_{\text{LP}}}_{i} \right) \mathbf{W}$$

$$+ \underbrace{\mathbf{N}_{\text{LP}} \cdots \mathbf{N}_{\text{LP}}}_{i} \hat{a}_0$$

$$= \left(\mathbf{I} + \mathbf{N}_{\text{LP}} + \mathbf{N}_{\text{LP}} \cdot \mathbf{N}_{\text{LP}} + \cdots + \underbrace{\mathbf{N}_{\text{LP}} \cdots \mathbf{N}_{\text{LP}}}_{i} \right) \mathbf{W} \qquad (27)$$

$$- \mathbf{W} + \underbrace{\mathbf{N}_{\text{LP}} \cdots \mathbf{N}_{\text{LP}}}_{i} \hat{a}_0.$$

In the case of $\text{Rank}(\mathbf{N}) = t$, $\mathbf{N}$ is the tth order positive definite matrix. An orthogonal matrix of order t is $\mathbf{Q}$ that follows the expression

$$\mathbf{N} + \lambda \mathbf{I} = \mathbf{Q}\,\mathbf{D}\mathbf{Q}^{\text{T}} + \lambda \mathbf{I} = \mathbf{Q}(\mathbf{D} + \lambda \mathbf{I})\mathbf{Q}^{\text{T}}, \qquad (28)$$

where $\mathbf{D} = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_t), \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_t > 0$, is the eigenvalue of $\mathbf{N}$.

Let $\mathbf{D} = \text{diag}(\sigma_1 + \lambda, \sigma_2 + \lambda, \ldots, \sigma_t + \lambda)$, then

$$\mathbf{N}_{\text{LP}} = (\mathbf{N} + \lambda \mathbf{I})^{-1} = \mathbf{Q}(\mathbf{D} + \lambda \mathbf{I})^{-1}\mathbf{Q}^{\text{T}} = \mathbf{Q}\mathbf{D}_1\mathbf{Q}^{\text{T}},$$

$$\|\mathbf{N}_{\text{LP}}\|_2^2 = \max \sigma_{\mathbf{N}_{\text{LP}}^{\text{T}}\mathbf{N}_{\text{LP}}}^{1/2} = \max_{1 \leq l \leq t} \left\{ (\sigma_l + 1)^{-1} \right\} = (\sigma_t + 1)^{-1} < 1. \qquad (29)$$

According to the convergence of the matrix sequence, it can be known that, in the case of $\|\mathbf{N}_{\text{LP}}\| < 1$, matrix $\underbrace{\mathbf{N}_{\text{LP}} \cdots \mathbf{N}_{\text{LP}}}_{i}$ is convergent and $\lim_{i \to \infty} \underbrace{\mathbf{N}_{\text{LP}} \cdots \mathbf{N}_{\text{LP}}}_{i} = 0$. Matrix power series $\sum_{i=0}^{\infty} \underbrace{\mathbf{N}_{\text{LP}} \cdots \mathbf{N}_{\text{LP}}}_{i}$ is absolutely convergent, and the sum is $(\mathbf{I} - \mathbf{N}_{\text{LP}})^{-1}$.

Therefore,

$$\lim_{i \to \infty} \hat{a}_i = \lim_{i \to \infty} \left[\left(\mathbf{I} + \mathbf{N}_{\text{LP}} + \mathbf{N}_{\text{LP}} \cdot \mathbf{N}_{\text{LP}} + \cdots + \underbrace{\mathbf{N}_{\text{LP}} \cdots \mathbf{N}_{\text{LP}}}_{i} \right) \mathbf{W} - \mathbf{W} \right] + \lim_{i \to \infty} \underbrace{\mathbf{N}_{\text{LP}} \cdots \mathbf{N}_{\text{LP}}}_{i} \hat{a}_0$$

$$= \left(\mathbf{I} + \mathbf{N}_{\text{LP}} + \mathbf{N}_{\text{LP}} \cdot \mathbf{N}_{\text{LP}} + \cdots + \underbrace{\mathbf{N}_{\text{LP}} \cdots \mathbf{N}_{\text{LP}}}_{i} + \cdots \right) \mathbf{W} - \mathbf{W}$$

$$= \left[(\mathbf{I} - \mathbf{N}_{\text{LP}})^{-1} - \mathbf{I} \right] \mathbf{W} = \left[(\mathbf{I} - \mathbf{N}_{\text{LP}})^{-1} (\mathbf{I} - (\mathbf{I} - \mathbf{N}_{\text{LP}})) \right] \mathbf{W} \qquad (30)$$

$$= \left[\mathbf{N}_{\text{LP}}^{-1} (\mathbf{I} - \mathbf{N}_{\text{LP}}) \right]^{-1} \mathbf{W} = \left(\mathbf{N}_{\text{LP}}^{-1} - \mathbf{I} \right)^{-1} \mathbf{W}$$

$$= \mathbf{N}^{-1} \mathbf{W}.$$

That is, in the case of $\text{Rank}(\mathbf{N}) = t$, the spectral correction iterative method based on L-curve converges to the least squares solution for any initial value. Therefore, the relationship between the estimated result $\hat{a}_m$ and the iteration termination value $\hat{a}_j$ can be further rewritten as

$$\hat{a}_m = \underbrace{\mathbf{N}_{\text{LP}} \cdots \mathbf{N}_{\text{LP}}}_{m-j} \mathbf{N}^{-1} \mathbf{W}$$

$$- \left(\underbrace{\mathbf{N}_{\text{LP}} \cdots \mathbf{N}_{\text{LP}}}_{m-j+1} + \underbrace{\mathbf{N}_{\text{LP}} \cdots \mathbf{N}_{\text{LP}}}_{m-j+2} + \cdots + \mathbf{N}_{\text{LP}}^{-1} + \mathbf{I} \right) \mathbf{W}. \qquad (31)$$

In the case of $\text{Rank}(\mathbf{N}) < t$, the initial value of $\hat{a}_0$ is equal to 0 in the spectral correction iterative method. Further, $\lim_{i \to \infty} \hat{a}_i = \mathbf{N}^{-1} \mathbf{W}$.

Therefore, this method is convergent with the least squares method and is also feasible.

2.4. Summarization of the Algorithm.
According to the estimation methods of linear parameters and nonlinear parameters proposed in Sections 2.1–2.3, the separable nonlinear least squares solution method—in which nonlinear parameters in the nonlinear least squares problem with SVD variable projection separation are solved with the LM algorithm and linear parameters are solved directly with the least squares method—is referred to as $\mathbf{LM}_{\text{VP}} + \mathbf{LS}$. When the linear parameters are solved using the iteration method that corrects characteristic values based on the L-curve, the method is referred to as $\mathbf{LM}_{\text{VP}} + \mathbf{CCVL}$. The traditional LM method without separation of parameters is referred to as $\mathbf{LM}_{\text{unSep}}$. The steps for solving the separable nonlinear least squares problem are summarized as follows.

A flowchart of Algorithm 1 is presented in Figure 1.

3. Numerical Examples

In this study, we used three examples, the Gaussian function fitting model, the fractional fitting model, and decomposition of full-waveform LiDAR data, to verify the method proposed in Section 2 for solving separable nonlinear least squares problems. The results were then compared with the LM algorithm with unseparated parameters with respect to iteration times, function calculation times, and fitting accuracy. The experiments were performed using MATLAB 2016b on a 2.3-GHz desktop PC running Windows 10.

3.1. Example 1: Gaussian Function Fitting Model.
The Gaussian function fitting model often exists in the parameter estimation of signal processing. For the generalized Gaussian

Step 1: given initial values $\mathbf{a}_0 \in \mathfrak{R}^p$, $\mathbf{b}_0 \in \mathfrak{R}^q$, set $\mu_0 = 0.1$, $(\rho, \sigma) \in (0, 1)$, $k = 0$. ε is a very small positive number.
Step 2: calculate the matrix $\Phi(\mathbf{b}_k)$ of the nonlinear function and Jacobian matrix $\mathbf{J}_k$ of the objective function after VP, using (17).
Step 3: calculate the iteration step length and iteration direction using (15) and (16), and then update the nonlinear parameters using (14).
Step 4: if $\|\mathbf{g}_k\|_2 = \|\mathbf{J}_k \mathbf{F}_k\|_2 \le \varepsilon$, terminate the algorithm; otherwise, $k = k + 1$, go to Step 3.
Step 5: calculate linear parameters when $\Phi^{\mathrm{T}}(\mathbf{b})\Phi(\mathbf{b})$ is a nonsingular matrix. The least squares method is used to calculate this directly; when $\Phi^{\mathrm{T}}(\mathbf{b})\Phi(\mathbf{b})$ is singular or the condition number of $\Phi(\mathbf{b})$ exceeds 100, the solution is obtained using (23).

ALGORITHM 1: Parameter estimation algorithm for the separable nonlinear least squares problem.
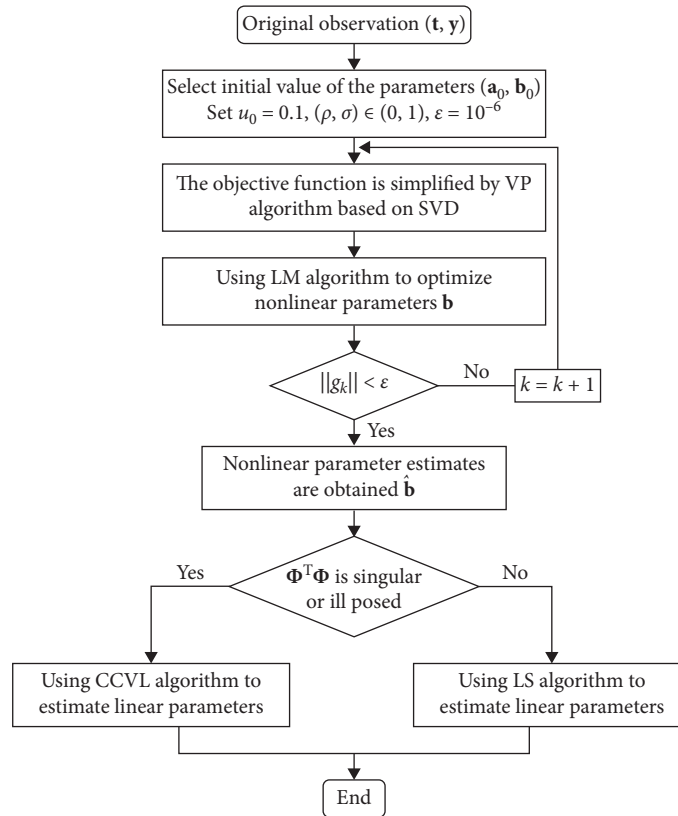

FIGURE 1: Flowchart for the parameter estimation process.

model, there is a linear combination of nonlinear functions that can be solved by separable nonlinear least squares based on the VP algorithm. The form of the model is as follows:

$$y(\mathbf{a}, \mathbf{b}; t_i) = \sum_{j=1}^{p} a_j \exp\left(\frac{-(t_i - u_j)^2}{2\theta_j^2}\right) + \xi_{t_i}, \quad i = 1, 2, \ldots, m,$$

$$(32)$$

where m is the number of observations, p is the number of linear parameters, and ξ_{t_i} is the Gaussian noise. $\mathbf{a} = (a_1, a_2, a_3, a_4, a_5, a_6)^{\mathrm{T}}$ is the linear parameter vector to be estimated. $\mathbf{b} = (u_1, \theta_1, u_2, \theta_2, u_3, \theta_3, u_4, \theta_4, u_5, \theta_5, u_6, \theta_6)^{\mathrm{T}}$

is the nonlinear parameter vector to be estimated. Equation (32) is written in the form of a matrix as

$$\mathbf{y}(\mathbf{a}, \mathbf{b}; \mathbf{t}) = \Phi_{m \times p}(\mathbf{u}, \boldsymbol{\theta}; \mathbf{t})\mathbf{a}_{p \times 1} + \boldsymbol{\xi}_{m \times 1}, \tag{33}$$

where $\Phi(\mathbf{u}, \boldsymbol{\theta}; \mathbf{t}) = [\phi_1, \phi_2, \ldots, \phi_p]$ and $\varphi_i = \exp(-(\mathbf{t}-u_i)^2/2\theta_i^2)$. The true parameter values of the model in (32) are $\mathbf{a}^* = (50, 40, 60, 60, 40, 50)^{\mathrm{T}}$, $\mathbf{u}^* = (50, 100, 150, 200, 250, 300)^{\mathrm{T}}$, and $\boldsymbol{\theta}^* = (10, 5, 15, 12, 20, 15)^{\mathrm{T}}$. The observation error $\boldsymbol{\xi}$ is a normal distribution with a standard deviation of 0.5. According to the above parameters, ten groups of observations are randomly generated, where each group has 400 data points. The initial values of the nonlinear and linear

parameters are randomly selected according to the following uniform distribution:

$$a_1^0 \sim U(45, 55),$$

$$u_1^0 \sim U(40, 60),$$

$$\theta_1^0 \sim U(8, 12),$$

$$a_2^0 \sim U(35, 45),$$

$$u_2^0 \sim U(90, 110),$$

$$\theta_2^0 \sim U(3, 8),$$

$$a_3^0 \sim U(55, 65),$$

$$u_3^0 \sim U(140, 160),$$

$$\theta_3^0 \sim U(10, 20),$$

$$a_4^0 \sim U(55, 65),$$

$$u_4^0 \sim U(180, 220),$$

$$\theta_4^0 \sim U(10, 15),$$

$$a_5^0 \sim U(35, 45),$$

$$u_5^0 \sim U(220, 280),$$

$$\theta_5^0 \sim U(15, 25),$$

$$a_6^0 \sim U(55, 65),$$

$$u_6^0 \sim U(280, 330),$$

$$\theta_6^0 \sim U(15, 20).$$

(34)

The error equations were listed based on the model in (32) and 400 pairs of observed values (t_i, y_i), and the parameters are solved using the $\mathbf{LM}_{\mathrm{unSep}}$, $\mathbf{LM}_{\mathrm{VP}} + \mathbf{LS}$, and $\mathbf{LM}_{\mathrm{VP}} + \mathbf{CCVL}$ methods, respectively. The parameter estimates are shown in Table 1, which are obtained by taking the average of 10 times.

The fitting curve is shown in Figure 2, and the difference between the parameter estimation and true value is shown in Figure 3. Because the estimated parameters obtained by the three methods do not differ significantly, and the fitting curves are basically coincident, only one fitting curve is drawn.

We can see from Table 1 that the results of the three methods are completely consistent in their estimation of the nonlinear parameters ($\widehat{\mu}$ and $\widehat{\theta}$). For the estimation of linear parameters, the results obtained by the $\mathbf{LM}_{\mathrm{unSep}}$ and $\mathbf{LM}_{\mathrm{VP}} + \mathbf{LS}$ methods are identical and the results obtained by the $\mathbf{LM}_{\mathrm{VP}} + \mathbf{CCVL}$ method deviate slightly from the true values compared to those obtained by the other methods. From Figure 2, we can see that the parameters estimated by the three methods can fit the observations very well and that their fitting curves are basically the same. Figure 3 shows that the difference value is closer to zero, and the parameter valuation is closer to the truth value. Therefore, the parameter estimation results of the $\mathbf{LM}_{\mathrm{unSep}}$ and $\mathbf{LM}_{\mathrm{VP}} + \mathbf{LS}$ methods may be taken to be identical and the difference value is small.

TABLE 1: Parameter estimation and true values of Example 1.

| Parameter	True value	$\mathbf{LM}_{\mathrm{unSep}}$	$\mathbf{LM}_{\mathrm{VP}} + \mathbf{LS}$	$\mathbf{LM}_{\mathrm{VP}} + \mathbf{CCVL}$
a_1	50	49.9938	49.9938	49.9435
a_2	40	40.1286	40.1286	40.1981
a_3	60	59.9751	59.9751	59.9469
a_4	60	60.0776	60.0776	60.0413
a_5	40	40.0097	40.0097	39.9811
a_6	50	50.0243	50.0243	50.3550
u_1	50	49.9994	49.9994	49.9994
θ_1	10	9.9845	9.9845	9.9845
u_2	100	99.9979	99.9979	99.9979
θ_2	5	4.9961	4.9961	4.9961
u_3	150	150.0055	150.0055	150.0055
θ_3	15	15.0132	15.0132	15.0132
u_4	200	200.0086	200.0086	200.0086
θ_4	12	11.9807	11.9807	11.9807
u_5	250	249.9897	249.9897	249.9897
u_6	300	300.0146	300.0146	300.0146
θ_6	15	14.9772	14.9772	14.9772
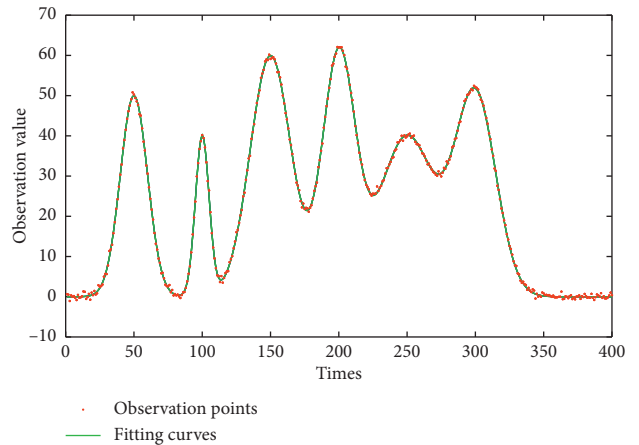
FIGURE 2: Observation points and fitting curves obtained for Example 1 for the three methods.
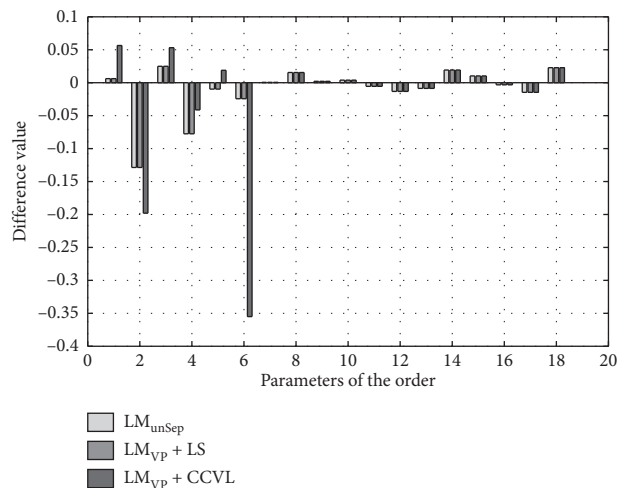

FIGURE 3: Differences between true value and estimation of parameters obtained by the three methods for Example 1.

To better understand the comparison of the parameter estimation results obtained using the three methods, Table 2 lists the maximum, minimum, and average values of the sum of squares of residuals between all the parameter estimation values and the true values (all_SSR), which were calculated ten times. Because the latter two methods eliminate linear parameters using the VP algorithm, the sum of squares of residuals between the nonlinear estimated parameters and the true value (nonpar_SSR) are also listed in Table 2.

We compare the fitting residual sum of squares and calculation process of the three methods in Table 3.

Table 2 indicates that, in terms of numerical value, the all_SSR values obtained by $LM_{VP} + LS$ and LM_{unSep} methods are equal. The results of nonpar_SSR using $LM_{VP} + CCVL$ are equal to those of the other two methods; however, the results of all_SSR are higher, indicating that the estimation of linear parameters is worse than that of the other two methods.

Table 3 indicates that the sum of squares of residual errors between the predicted and true values obtained using $LM_{VP} + CCVL$ is the largest, and those of the other two methods are equal. The mean square errors of the LM_{unSep}, $LM_{VP} + LS$, and $LM_{VP} + CCVL$ methods are 0.0107, 0.0107, and 0.0248, respectively, which are relatively small, indicating that the results obtained by all three methods are reliable. Compared with the number of iterations and average number of calculations of the function, the method based on VP exhibits a considerable reduction compared with the method in which the parameters are not separated.

The residual change of the objective function in the iterative process is shown in Figure 4.

According to Figure 4, in the case of Gaussian function fitting, on the whole, the residual changes of the three methods are all reduced and the estimated results of the final parameters are very close to the truth value, suggesting that the three methods are convergent.

In addition, according to the structural characteristics of the model, when the nonlinear parameters are fixed in the $LM_{VP} + LS$ and $LM_{VP} + CCVL$ methods, the nonlinear function matrix is full rank and the number of conditions is less than five for all ten calculations; thus, there is no ill-condition problem.

Based on the experimental results, we draw two conclusions: (1) Because there is no rank deficiency or ill posing in the matrix of the nonlinear function, the LS method helps to attain more accurate linear parameters directly than the regularization method. (2) The $LM_{VP} + LS$ method eliminates the linear parameters using the VP algorithm based on SVD, reduces the dimension of the parameters to be estimated, and improves the efficiency of the iteration process. The number of iterations required in the $LM_{VP} + LS$ method is less than that for the LM_{unSep} method and produces the best effect under the same precision.

3.2. Example 2: Fractional Fitting Model.
The fractional fitting model is also a commonly used curve-fitting method. The model is described by the following expression:

$$y(\mathbf{a}, \mathbf{b}; t) = a_1 + a_2 t + a_3 t^2$$

$$-a_4 \left[\frac{1}{1 + ((b_1 + 0.5b_2 - t)/b_3)^2} + \frac{1}{1 + ((b_1 - 0.5b_2 - t)/b_3)^2} \right]$$

$$-a_5 \left[\frac{1}{1 + ((b_4 + 0.5b_5 - t)/b_6)^2} + \frac{1}{1 + ((b_4 - 0.5b_5 - t)/b_6)^2} \right] \tag{35}$$

$$-a_6 \left[\frac{1}{1 + ((b_7 - t)/b_8)^2} \right] + \xi_t,$$

where $\mathbf{a} = (a_1, a_2, a_3, a_4, a_5, a_6)^T$ is a linear parameter vector to be estimated and $\mathbf{b} = (b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8)^T$ is the nonlinear parameter vector. The model in (35) is written in the form of a matrix as follows:

$$\mathbf{y}(\mathbf{a}, \mathbf{b}; \mathbf{t}) = \mathbf{\Phi}_{200 \times 6}(\mathbf{b}; \mathbf{t}) \mathbf{a}_{6 \times 1} + \mathbf{\xi}_{200 \times 1}, \tag{36}$$

where $\mathbf{\Phi}(\mathbf{u}, \mathbf{\theta}; \mathbf{t}) = [\varphi_1, \varphi_2, \ldots, \varphi_6]$ and φ_i $(i = 1, 2, \ldots, 6)$ is the coefficient term corresponding to the linear parameter a_i. The true values of the parameters in the model in (35) are $\mathbf{a}^* = (5.0, 0.5, 0.3, 0.4, 0.5, 0.7)^T$ and $\mathbf{b}^* = (0.5, 1, 2, 0.8, 0.8, 2.5, 0.8, 3)^T$. Observation error $\mathbf{\xi}$ is a normal distribution with a standard deviation of 0.001. According to the above

parameters, ten groups of observations are randomly generated, where each group has 200 data points. The initial values of the nonlinear and linear parameters are randomly selected according to the following uniform distribution:

$$\mathbf{a}^0 = \mathbf{a}^* + \xi(a),$$
$$\xi(a) \sim N(0, 0.01),$$
$$\mathbf{b}^0 = \mathbf{b}^* + \xi(b), \tag{37}$$
$$\xi(b) \sim N(0, 0.01).$$

The error equations listed are based on the model in (32) and 200 pairs of observed values (t_i, y_i), and the parameters

TABLE 2: Sum of the squared residuals between the estimated parameters obtained using the three methods and the true values in Example 1.

Method	all_SSR (max)	all_SSR (min)	all_SSR (mean)	nonpar_SSR (max)	nonpar_SSR (min)	nonpar_SSR (mean)
$\mathbf{LM_{unSep}}$	0.2930	0.0371	0.1715	0.1699	0.0113	0.0424
$\mathbf{LM_{VP} + LS}$	0.2930	0.0371	0.1715	0.1699	0.0113	0.0424
$\mathbf{LM_{VP} + CCVL}$	0.9767	0.1864	0.5134	0.1699	0.0113	0.0424

TABLE 3: Sum of the squared residuals between predicted values obtained using the three methods and the true observations in Example 1.

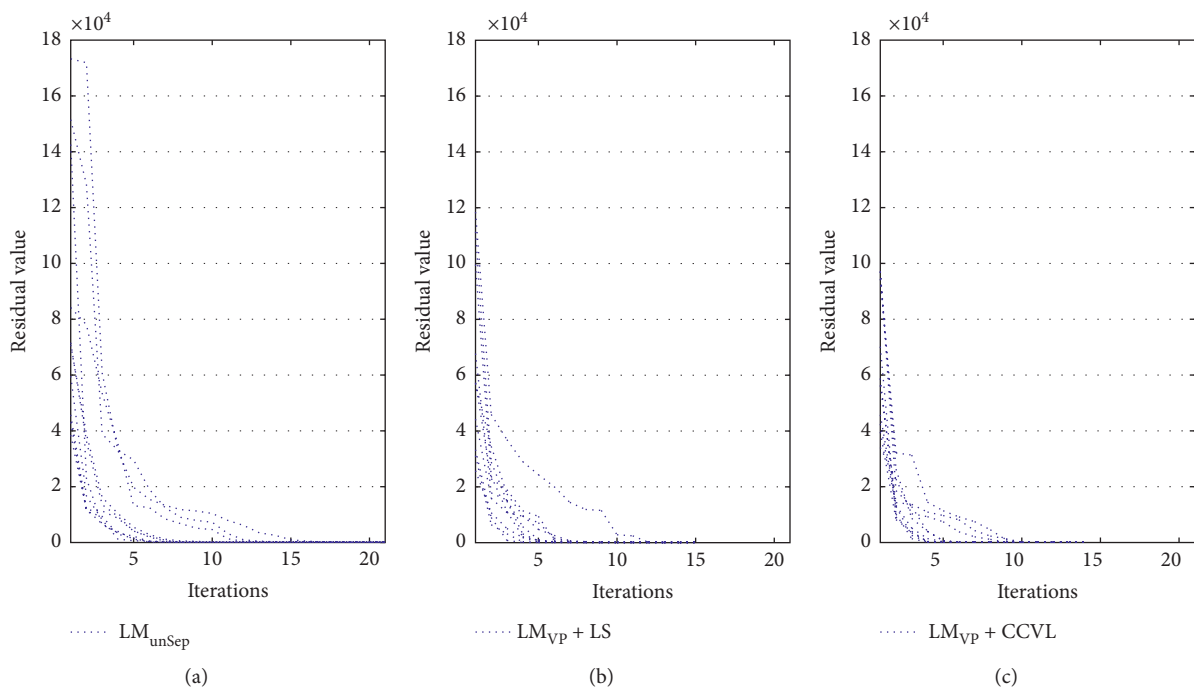| Method	y_SSR (max)	y_SSR (min)	y_SSR (mean)	MSE	Iteration times (mean)	Function count (mean)
$\mathbf{LM_{unSep}}$	5.0763	3.0089	4.2680	0.0107	11.9	251.5
$\mathbf{LM_{VP} + LS}$	5.0763	3.0089	4.2680	0.0107	9.8	144.9
$\mathbf{LM_{VP} + CVCL}$	14.3523	7.3535	9.9158	0.0248	9.8	144.9

FIGURE 4: Residual changes of the three methods in the iteration process for Example 1.

are solved using the $\mathbf{LM_{unSep}}$, $\mathbf{LM_{VP} + LS}$, and $\mathbf{LM_{VP} + CCVL}$ methods. The parameter estimates obtained by taking the average of the ten calculations are listed in Table 4.

Because the estimated parameters obtained by the three methods do not differ significantly, and the fitting curves are basically coincident, only one fitting curve is drawn. The fitting curve is shown in Figure 5.

Table 4 and Figure 5 indicate that the three methods produce similar true values for the nonlinear parameters. The calculation results in the $\mathbf{LM_{VP} + CCVL}$ and $\mathbf{LM_{VP} + LS}$ methods are the same because the separation of the objective function is equal, and the nonlinear parameters are all estimated by the LM algorithm. When the $\mathbf{LM_{VP} + LS}$ method is used, significant deviation is observed between the true values of the linear parameters. Combined with Figure 5, the fitting curves of the three methods are similar, which shows that the $\mathbf{LM_{VP} + LS}$ method cannot obtain the optimal parameters even though it satisfies the least squares principle in solving the linear parameters. This is because the condition number of the nonlinear function matrix is considerably greater than 100 in the process of solving the separation problem after VP. The number of conditions is above 10^7 for all ten calculations, and these were seriously ill-posed problems.

Because the differences in the linear parameters solved by the $\mathbf{LM_{VP} + LS}$ method are much higher than those of the nonlinear parameters, the original difference was reduced by 10^4 and then plotted. The differences between the parameter estimation solutions of the three methods and the true value are shown in Figure 6.

Figure 6 indicates that the difference between the parameter estimation and the true value obtained by the $\mathbf{LM_{VP} + CCVL}$ method is the smallest among the three methods, that is, the result is closest to the true values. The

Table 4: Parameters estimation and true values for Example 2.

| Order	Parameter	True value	LM_{unSep}	$LM_{VP} + LS$	$LM_{VP} + CCVL$
1	$\widehat{a}_1$	5	4.9631	−14618.7461	4.9955
2	$\widehat{a}_2$	0.5	0.7870	−724.1217	0.5343
3	$\widehat{a}_3$	0.3	0.2143	672.7237	0.2999
4	$\widehat{a}_4$	0.4	0.4850	284.4777	0.3869
5	$\widehat{a}_5$	0.5	0.6116	13585.6419	0.5080
6	$\widehat{a}_6$	0.7	0.6835	−41575.4314	0.7272
7	$\widehat{b}_1$	0.5	0.5356	0.5784	0.5784
8	$\widehat{b}_2$	1	0.9094	0.9502	0.9502
9	$\widehat{b}_3$	2	1.2188	2.0133	2.0133
10	$\widehat{b}_4$	0.8	0.9804	0.8459	0.8459
11	$\widehat{b}_5$	0.8	0.6996	0.7919	0.7919
12	$\widehat{b}_6$	2.5	2.2681	2.5049	2.5049
13	$\widehat{b}_7$	0.8	0.8597	0.8080	0.8080
14	$\widehat{b}_8$	3	3.0277	3.0042	3.0042
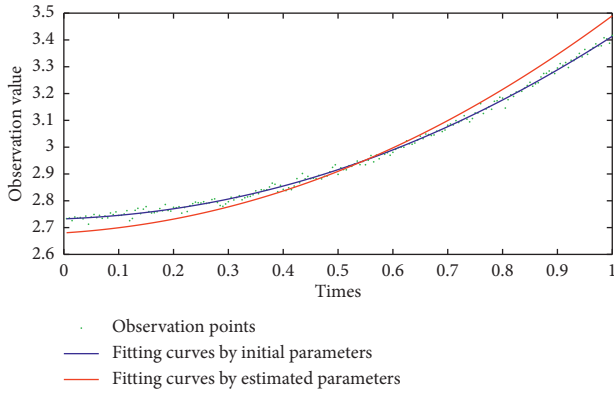
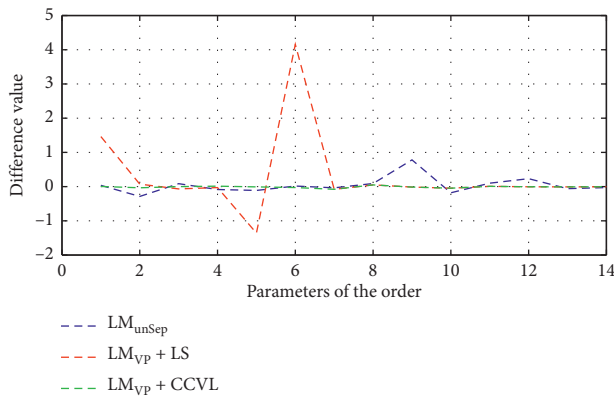Figure 5: Observation points and fitting curves obtained using the three methods for Example 2.

Figure 6: Difference between the true value and the estimation of parameters obtained using the three methods in Example 2.

difference obtained by the LM_{unSep} method is next. The difference obtained by the $LM_{VP} + LS$ method is the largest, even if that is reduced by 10^4.

To better understand the comparison of the parameter estimation results of the three methods, Table 5 lists the maximum, minimum, and average values of the sum of squares of residuals between all the parameter estimation values and the true values (all_SSR) obtained from the ten calculations. Because the latter two methods eliminate linear parameters using the VP algorithm, the sums of squares of residuals between nonlinear estimated parameters and the true value (nonpar_SSR) are also listed in Table 5.

Table 5 indicates that the sum of the residuals of all the parameters using the $LM_{VP} + LS$ method is much larger than the other two methods. The main reason is that the matrix of the nonlinear function is seriously ill posed, and the results of direct solution using the least squares method are not the optimal parameters. Therefore, the $LM_{VP} + CCVL$ method using the iteration method by correcting characteristic values based on the L-curve obtained the optimal estimates that are similar to the true values. The sum of the residual squares of parameter estimates and true values is the smallest. The overall result is also superior to the LM_{unSep} method when the parameters are not separated. For the sum of residual squares of nonlinear parameters, the $LM_{VP} + CCVL$ and $LM_{VP} + LS$ methods exhibit the same results, which are lower than the values obtained by the LM_{unSep} method.

A comparison of the three methods in terms of model fitting and calculation processes is presented in Table 6.

Table 6 indicates that in terms of the sum of squared residues between the predicted value and the true value, irrespective of the maximum, minimum, or average value, the lowest values are obtained when the $LM_{VP} + CCVL$ method is used, followed by the LM_{unSep} method; highest values are obtained using the $LM_{VP} + LS$ method. In terms of the mean square error, the $LM_{VP} + CCVL$ method provides the lowest values. The number of iterations and average number of functions considerably decrease for the VP algorithm; the average number of iterations was reduced from 161.4 to 20.8 and the average calculation number of functions was reduced from 2516.2 to 207.7.

The residual change in the objective function of the iterative process is shown in Figure 7.

According to Figure 7, in the case of Gaussian function fitting, on the whole, the residual changes of the three methods are all reduced and the estimated results of the final parameters are very close to the truth value, so it can be seen that the three methods are convergent.

The experimental analysis indicates that the problem is seriously ill posed, and when considering separable nonlinear least squares, the results obtained by the $LM_{VP} + CCVL$ method afford the highest fitting precision. The computational efficiency is also much better than that for the LM method without separation.

3.3. Example 3: Decomposition of Full-Waveform LiDAR Data.
This section of the experimental data originated from the full-waveform LiDAR data of a region collected in 2016. The full-waveform lidar system record the backscattered

TABLE 5: Sums of the squared residuals between the estimated parameters obtained using the three methods and the true values in Example 2.

Method	all_SSR (max)	all_SSR (min)	all_SSR (mean)	nonpar_SSR (max)	nonpar_SSR (min)	nonpar_SSR (mean)
$\mathbf{LM}_{\text{unSep}}$	5.4669	0.9927	2.4807	0.0227	0.0159	0.01915
$\mathbf{LM}_{\text{VP}} + \mathbf{LS}$	$1.44E + 11$	$6.59E + 08$	$2.58E + 10$	0.0226	0.0155	0.01896
$\mathbf{LM}_{\text{VP}} + \mathbf{CCVL}$	0.307	0.0168	0.1232	0.0226	0.0155	0.01896

TABLE 6: Sums of the squared residuals between predicted value obtained by the three methods and the true observations in Example 2.

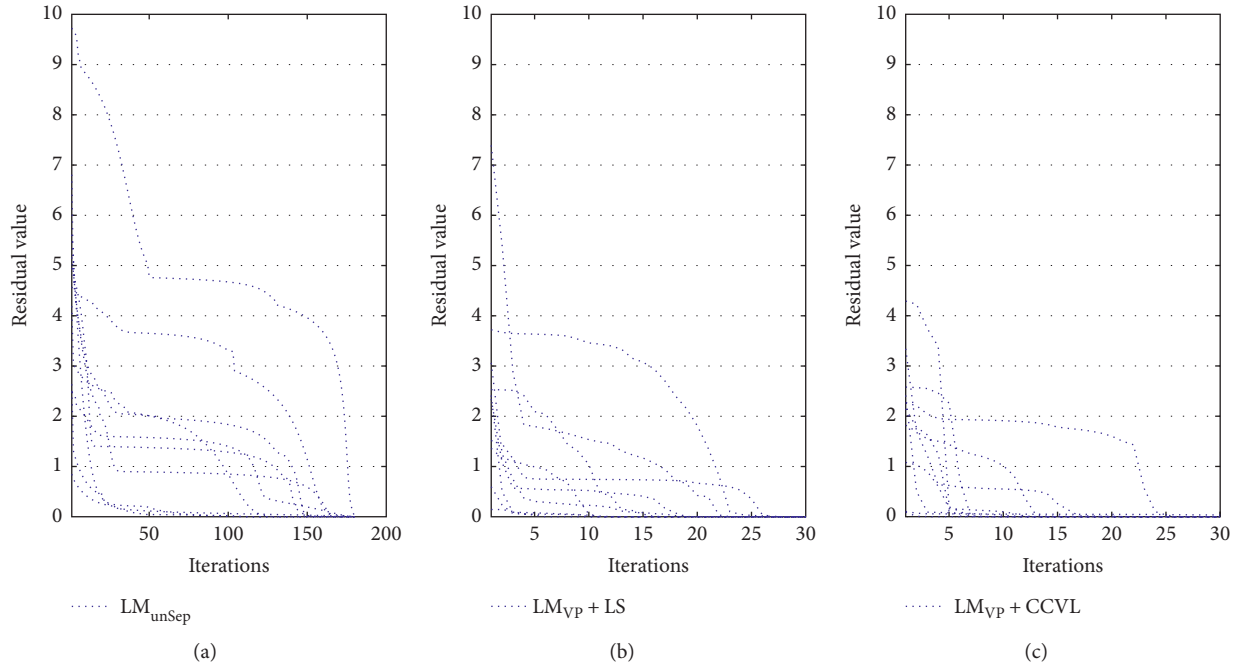| Method	y_SSR (max)	y_SSR (min)	y_SSR (mean)	MSE	Iteration times (mean)	Function count (mean)
$\mathbf{LM}_{\text{unSep}}$	$7.30E - 04$	$2.60E - 04$	$3.78E - 04$	$1.89E - 06$	161.4	2516.2
$\mathbf{LM}_{\text{VP}} + \mathbf{LS}$	$8.37E - 04$	$3.60E - 04$	$5.84E - 04$	$2.92E - 06$	20.8	207.7
$\mathbf{LM}_{\text{VP}} + \mathbf{CCVL}$	$4.12E - 04$	$3.17E - 05$	$1.90E - 04$	$9.52E - 07$	20.8	207.7

FIGURE 7: Residual changes of the three methods in the iteration process of Example 2.

energy at different elevation points in a certain elevation range in the form of waveforms. The survey area mainly contains buildings, trees, and roads. The point cloud and waveform information are stored in the LAS file in the LAS1.3 standard format. The waveform data sampling interval is 1 ns and the number of samples is 287. The echo waveform is regarded as the superposition of several Gaussian functions. The model form is as follows:

$$y(\mathbf{a}, \mathbf{b}; t_i) = \sum_{j=1}^{4} a_j \exp\left(\frac{-(t_i - u_j)^2}{2\theta_j^2}\right) + \xi_{t_i}, \quad i = 1, 2, \ldots, 287,$$
(38)

where ξ_{t_i} is the Gaussian noise, $\mathbf{a} = (a_1, a_2, a_3, a_4)^{\text{T}}$ is the linear parameter vector to be estimated, and $\mathbf{b} = (\mu_1, \theta_1,$

$\mu_2, \theta_2, \mu_3, \theta_3, \mu_4, \theta_4)^{\text{T}}$ is the nonlinear parameter vector to be estimated.

The parameters were solved by the $\mathbf{LM}_{\text{unSep}}$, $\mathbf{LM}_{\text{VP}} + \mathbf{LS}$, and $\mathbf{LM}_{\text{VP}} + \mathbf{CCVL}$ methods. The fitting curves are shown in Figures 8(a)–8(c).

It can be seen from Figure 8 that, in the case of the $\mathbf{LM}_{\text{unSep}}$ method, only the first decomposition waveform can be fitted to the observation value and the latter three decomposition waveforms are not well fitted. $\mathbf{LM}_{\text{VP}} + \mathbf{LS}$ and $\mathbf{LM}_{\text{VP}} + \mathbf{CCVL}$ provide an accurate fitting result.

As there is no truth value in the actual measurement, the estimated result is no longer compared with the truth value. To better understand the comparison of the parameter estimation results of the three methods, the mean square error between the fitting results and the observed values, the maximum fit difference (Diff-max), the minimum fit
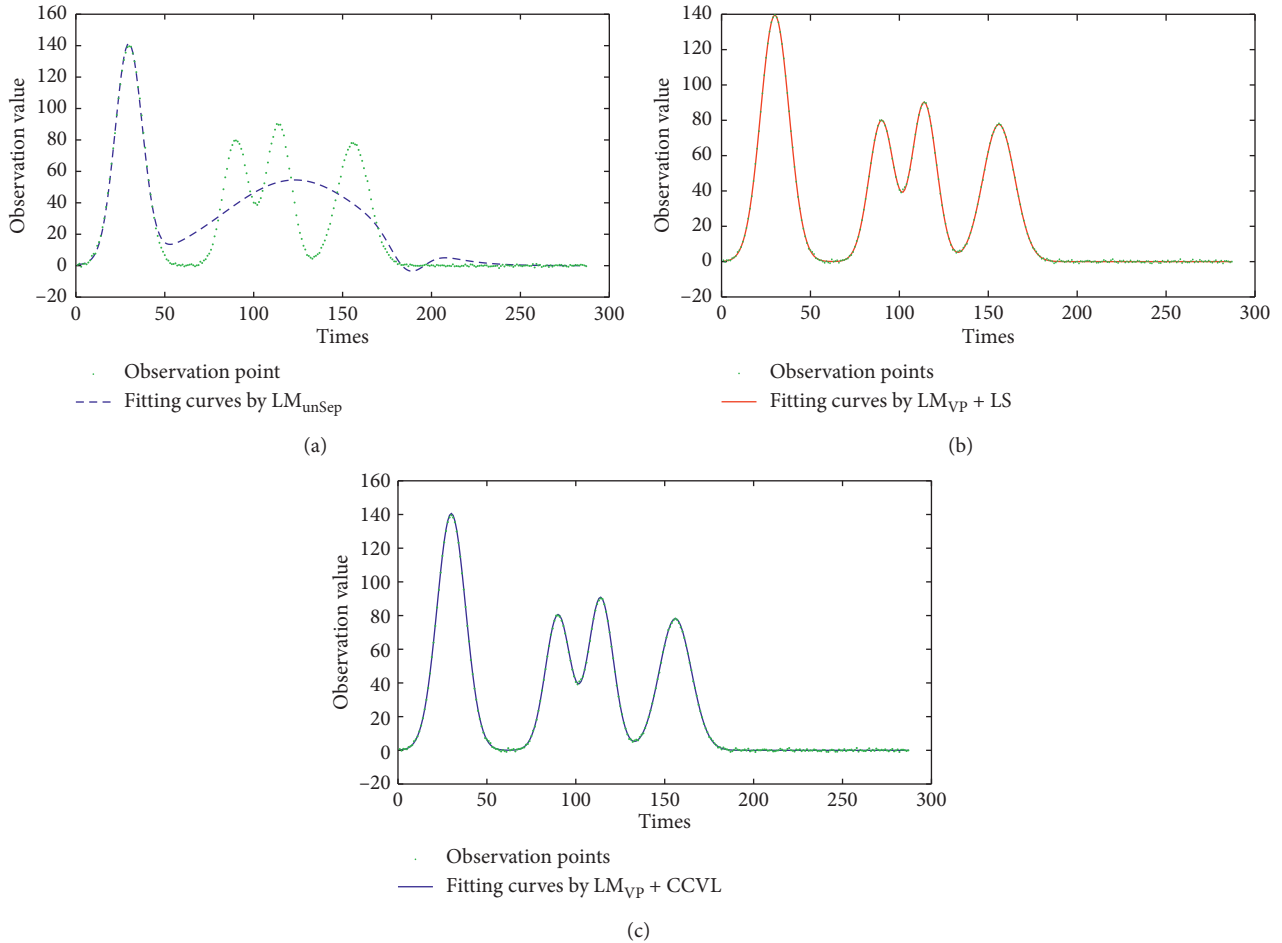
(a)

(b)

(c)

FIGURE 8: Fitting curves of the $\mathbf{LM}_{unSep}$, $\mathbf{LM}_{VP} + \mathbf{LS}$, and $\mathbf{LM}_{VP} + \mathbf{CCVL}$ methods.

TABLE 7: Comparison of the solution results of the three methods.

Method	Diff-max	Diff-min	MSE	Iteration times	Function count
$\mathbf{LM}_{unSep}$	48.0734	0.0191	285.8365	20	290
$\mathbf{LM}_{VP} + \mathbf{LS}$	1.4111	0.0076	0.2577	10	106
$\mathbf{LM}_{VP} + \mathbf{CCVL}$	1.8342	0.0004	0.2577	10	106

difference (Diff-min), the number of iterations, and function count calculations of the three methods are presented in Table 7.

It can be seen from Table 7 that, like the simulation experiment results, owing to parameter separation, the reduced dimension of the parameter estimation improves the chance of convergence. $\mathbf{LM}_{VP} + \mathbf{LS}$ is the best of the three methods, with the least number of iterations and function calculations. $\mathbf{LM}_{VP} + \mathbf{CCVL}$ is second. However, the $\mathbf{LM}_{unSep}$ method, in which parameters are not separated, does not produce the correct result.

The residual changes of the objective function in the iterative process are shown in Figure 9.

From Figure 9, we can know that all three methods tend to converge, but $\mathbf{LM}_{unSep}$, whose parameters are not separated, does not converge to the optimal result, and the sum of squared residual convolutions is still large. The sum of the residuals of the remaining two methods tends to zero. The number of iterations is reduced from 20 to 10, and function count is also decreased from 290 to 106. The calculation is greatly simplified.

In short, these experiments indicate that the estimation method of parameter separation shows a great improvement in terms of the number of iterations, number of function calculations, and calculation result.

4. Conclusion

In this study, linear and nonlinear parameters were separated by the VP algorithm based on SVD. Further, the separable least squares problem was transformed into a least squares problem with only nonlinear parameters, which reduced the dimension of the parameters, number of iterations, and function calculations and improved operation
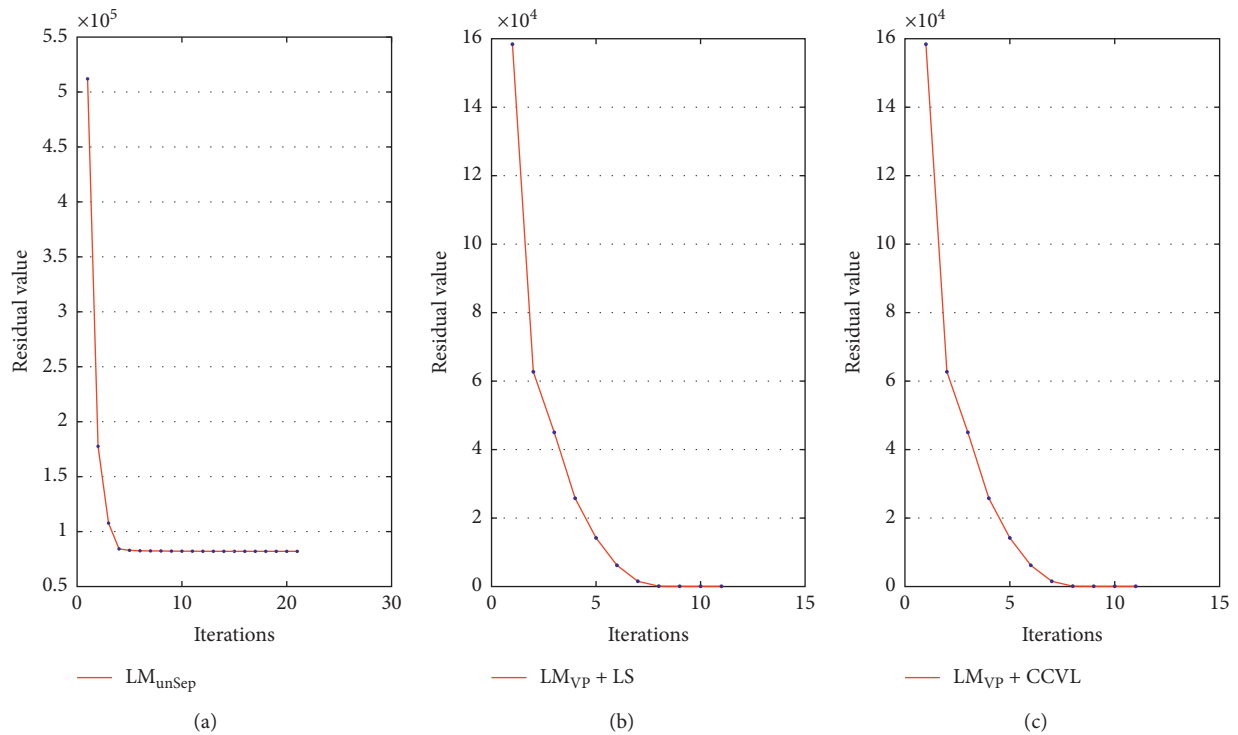
FIGURE 9: Residual changes of the **LM**$_{unSep}$, **LM**$_{VP}$ + **LS**, and **LM**$_{VP}$ + **CCVL** methods in the iteration process of Example 3.

efficiency. In addition, to solve the ill-posed problem of the coefficient matrix comprising nonlinear functions when solving for linear parameters, an iteration method that corrects characteristic values based on the L-curve was adopted. This also helped to ensure the convergence of model parameter estimation and improved prediction accuracy. The parameter estimation method used in our study is suitable for cases with a large number of linear parameters in separable nonlinear least squares problems. Additionally, an iteration method that involves correcting characteristic values based on L-curve was used to solve the ill-conditioned coefficient matrix while solving linear parameters. However, one limitation is that the situation of rank deficit often occurred in the process of separable nonlinear least squares parameter estimation; this will be addressed in future research.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] G. H. Golub and V. Pereyra, "The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate," *SIAM Journal on Numerical Analysis*, vol. 10, no. 2, pp. 413–432, 1973.

[2] R. Savitha, S. Suresh, and N. Sundararajan, "Metacognitive learning in a fully complex-valued radial basis function neural network," *Neural Computation*, vol. 24, no. 5, pp. 1297–1328, 2012.

[3] R. Rakkiyappan, A. Chandrasekar, S. Laksmanan, and J. H. Park, "State estimation of memristor-based recurrent neural networks with time-varying delays based on passivity theory," *Complexity*, vol. 19, no. 4, pp. 32–43, 2014.

[4] K. B. Cho and B. H. Wang, "Radial basis function based adaptive fuzzy systems and their applications to system identification and prediction," *Fuzzy Sets and Systems*, vol. 83, no. 3, pp. 325–339, 1996.

[5] H.-X. Li and Z. Liu, "A probabilistic neural-fuzzy learning system for stochastic modeling," *IEEE Transactions on Fuzzy Systems*, vol. 16, no. 4, pp. 898–908, 2008.

[6] M. R. Osborne and G. K. Smyth, "A modified prony algorithm for exponential function fitting," *SIAM Journal on Scientific Computing*, vol. 16, no. 1, pp. 119–138, 1995.

[7] B. D. Rao and K. S. Arun, "Model based processing of signals: a state space approach," *Proceedings of the IEEE*, vol. 80, no. 2, pp. 283–309, 1992.

[8] P. C. Li, *The Technology of Terrain and Building Reconstruction Using Airborne Full-Waveform Li DAR Data*, Information Engineering University, Zhengzhou, China, 2015.

[9] U. Prells and M. I. Friswell, "Application of the variable projection method for updating models of mechanical systems," *Journal of Sound and Vibration*, vol. 225, no. 2, pp. 307–325, 1999.

[10] R. Roy and T. Kailath, "Esprit-estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 7, pp. 984–995, 1989.

[11] R. J. Gardner and P. Milanfar, "Reconstruction of convex bodies from brightness functions," *Discrete and Computational Geometry*, vol. 29, no. 2, pp. 279–303, 2003.

[12] G. Golub and V. Pereyra, "Separable nonlinear least squares: the variable projection method and its applications," *Inverse Problems*, vol. 19, no. 2, pp. R1–R26, 2003.

[13] J. Sjoberg and M. Viberg, "Separable non-linear least-squares minimization—possible improvements for neural net fitting," in *Proceedings of the 1997 IEEE Signal Processing Society Workshop. Neural Networks for Signal Processing VII*, Amelia Island, FL, USA, September 1997.

[14] Z. Q. Fu and L. L. Guo, "Tikhonov regularized variable projection algorithms for separable nonlinear least squares problems," *Complexity*, vol. 2019, Article ID 4861708, 9 pages, 2019.

[15] L. Kaufman, "A variable projection method for solving separable nonlinear least squares problems," *BIT*, vol. 15, no. 1, pp. 49–57, 1975.

[16] A. E. B. Ruano, D. I. Jones, and P. J. Fleming, "A new formulation of the learning problem of a neural network controller," in *Proceedings of the 30th IEEE Conference on Decision and Control*, pp. 865-866, Brighton, UK, December 1991.

[17] M. Gan and H.-X. Li, "An efficient variable projection formulation for separable nonlinear least squares problems," *IEEE Trans. Cybern*, vol. 44, no. 5, pp. 707–711, 2014.

[18] Z. Bai and W. Sun, "Existence and multiplicity of positive solutions for singular fractional boundary value problems," *Computers & Mathematics with Applications*, vol. 63, no. 9, pp. 1369–1381, 2012.

[19] C. Han, T. Feng, G. He, and T. Guo, "Parallel variable distribution algorithm for constrained optimization with nonmonotone technique," *Journal of Applied Mathematics*, vol. 2013, Article ID 295147, 7 pages, 2013.

[20] C. Y. Liu, R. Loxton, and K. L. Teo, "Optimal parameter selection for nonlinear multistage systems with time-delays," *Computational Optimization and Applications*, vol. 59, no. 1-2, pp. 285–306, 2014.

[21] Z. Fu, G. Liu, and L. Guo, "Sequential quadratic programming method for nonlinear least squares estimation and its application," *Mathematical Problems in Engineering*, vol. 2019, Article ID 3087949, 8 pages, 2019.

[22] Y. Cheng, Q. Mou, X. Pan, and S. Yao, "A sufficient descent conjugate gradient method and its global convergence," *Optimization Methods and Software*, vol. 31, no. 3, pp. 577–590, 2016.

[23] K. Madsen, H. B. Nielsen, and O. Tingleff, "Methods for nonlinear least squares problems," *Society for Industrial and Applied Mathematics*, vol. 2012, no. 1, pp. 1409–1415, 2004.

[24] K. Levenberg, "A method for the solution of certain nonlinear problems in least squares," *Quarterly of Applied Mathematics*, vol. 2, no. 4, pp. 164–168, 1944.

[25] D. W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *Journal of the Society for Industrial and Applied Mathematics*, vol. 11, no. 2, pp. 431–441, 1963.

[26] Z. Gong, C. Liu, K. L. Teo, and J. Sun, "Distributionally robust parameter identification of a time-delay dynamical system with stochastic measurements," *Applied Mathematical Modelling*, vol. 69, pp. 685–695, 2019.

[27] C. Liu, M. Han, Z. Gong, and K. L. Teo, "Robust parameter estimation for constrained time-delay systems with inexact measurements," *Journal of Industrial & Management Optimization*, vol. 13, no. 5, 2017.

[28] P. Torres, J.-W. van Wingerden, and M. Verhaegen, "Output-error identification of large scale 1D-spatially varying interconnected systems," *IEEE Transactions on Automatic Control*, vol. 60, no. 1, pp. 130–142, 2015.

[29] X. Y. Dong, Z. B. Bai, and S. Q. Zhang, "Positive solutions to boundary value problems of p-Laplacian with fractional derivative," *Boundary Value Problems*, vol. 2017, no. 1, pp. 1–15, 2017.

[30] T. Ma, X. Meng, and Z. Chang, "Dynamics and optimal harvesting control for a stochastic one-predator-two-prey time delay system with jumps," *Complexity*, vol. 2019, Article ID 5342031, 19 pages, 2019.

[31] S. Bellavia, S. Gratton, and E. Riccietti, "A Levenberg-Marquardt method for large nonlinear least-squares problems with dynamic accuracy in functions and gradients," *Numerische Mathematik*, vol. 140, no. 3, pp. 791–825, 2018.

[32] T. Jiang, Z. Jiang, and S. Ling, "An algebraic method for quaternion and complex Least Squares coneigen-problem in quantum mechanics," *Applied Mathematics and Computation*, vol. 249, pp. 222–228, 2014.

[33] C. Liu, Z. Gong, and K. L. Teo, "Robust parameter estimation for nonlinear multistage time-delay systems with noisy measurement data," *Applied Mathematical Modelling*, vol. 53, pp. 353–368, 2018.

[34] F. Ding, L. Xu, F. E. Alsaadi, and T. Hayat, "Iterative parameter identification for pseudo-linear systems with ARMA noise using the filtering technique," *IET Control Theory & Applications*, vol. 12, no. 7, pp. 892–899, 2018.

[35] F. Ding, F. Wang, L. Xu, and M. Wu, "Decomposition based least squares iterative identification algorithm for multivariate pseudo-linear ARMA systems using the data filtering," *Journal of the Franklin Institute*, vol. 354, no. 3, pp. 1321–1339, 2017.

[36] A. Gholami and H. R. Siahkoohi, "Regularization of linear and non-linear geophysical ill-posed problems with joint sparsity constraints," *Geophysical Journal International*, vol. 180, no. 2, pp. 871–882, 2010.

[37] J. Lampe and H. Voss, "Large-scale Tikhonov regularization of total least squares," *Journal of Computational and Applied Mathematics*, vol. 238, pp. 95–108, 2012.

[38] P. C. Hansen, "Truncated singular value decomposition solutions to discrete III-posed problems with III-determined numerical rank," *SIAM Journal on Scientific and Statistical Computing*, vol. 11, no. 3, pp. 503–518, 1990.

[39] X. Z. Wang, D. Y. Liu, Q. Y. Zhang, and H. L. Huang, "The iteration by correcting characteristic value and its application in surveying data processing," *Journal of Heilongjiang Institute of Technology*, vol. 15, no. 2, pp. 3–6, 2001.

[40] X. Z. Wang, D. Y. Liu, and H. L. Huang, "The co-factor matrix of the iteration method by correcting characteristic value," *Geomatics and Information Science of Wuhan University*, vol. 28, no. 4, pp. 429–431, 2003.

[41] G. Y. Chen, M. Gan, C. L. P. Chen, and H. X. Li, "A regularized variable projection algorithm for separable nonlinear least squares problems," *IEEE Transactions on Automatic Control*, vol. 64, no. 2, pp. 526–537, 2019.

[42] M. Zhai, G. Liu, Q. Tao, and M. Xin, "A novel characteristic value correction iteration method," *Communications in Statistics—Simulation and Computation*, vol. 48, no. 2, pp. 591–600, 2019.

[43] P. C. Hansen, "Analysis of discrete III-posed problems by means of the L-curve," *SIAM Review*, vol. 34, no. 4, pp. 561–580, 1992.

[44] S. Oh and S. Kwon, "Choosing regularization parameter by global L-curve criterion," *Journal of Chungcheong Mathematical Society*, vol. 30, no. 1, pp. 117–128, 2017.

[45] R. H. Byrd, R. B. Schnabel, and G. A. Shultz, "A trust region algorithm for nonlinearly constrained optimization," *SIAM Journal on Numerical Analysis*, vol. 24, no. 5, pp. 1152–1170, 1987.