

Article

An Improved Tikhonov-Regularized Variable Projection Algorithm for Separable Nonlinear Least Squares

Hua Guo ^{1,2,*} , Guolin Liu ¹ and Luyao Wang ¹

¹ College of Geodesy and Geomatics, Shandong University of Science and Technology, Qingdao 266590, China; gliu@sdust.edu.cn (G.L.); wly_199498@163.com (L.W.)

² College of Mathematics and Systems Science, Shandong University of Science and Technology, Qingdao 266590, China

* Correspondence: skd993121@sdust.edu.cn; Tel.: +86-138-6429-1557

Abstract: In this work, we investigate the ill-conditioned problem of a separable, nonlinear least squares model by using the variable projection method. Based on the truncated singular value decomposition method and the Tikhonov regularization method, we propose an improved Tikhonov regularization method, which neither discards small singular values, nor treats all singular value corrections. By fitting the Mackey–Glass time series in an exponential model, we compare the three regularization methods, and the numerically simulated results indicate that the improved regularization method is more effective at reducing the mean square error of the solution and increasing the accuracy of unknowns.

Keywords: separable nonlinear least squares; variable projection method; Tikhonov regularization method; L-curve method; LM algorithm



Citation: Guo, H.; Liu, G.; Wang, L. An Improved Tikhonov-Regularized Variable Projection Algorithm for Separable Nonlinear Least Squares. *Axioms* **2021**, *10*, 196. <https://doi.org/10.3390/axioms10030196>

Academic Editor: Hijaz Ahmad

Received: 17 July 2021

Accepted: 19 August 2021

Published: 22 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Many problems in physics, chemistry, machine learning, computer vision, signal processing and mechanical engineering can only be described by a specialized type of nonlinear regression model, which is a linear combination of nonlinear functions. In particular, given the sequence of the observed values $\{(t_i, y_i), i = 1, 2, \dots, m\}$, a fitting model can be established as follows:

$$\eta(\theta_L, \theta_N; t_i) := \sum_{j=1}^n a_j \varphi_j(\theta_N; t_i) \quad (1)$$

where $\theta_L = (a_1, a_2, \dots, a_n)^T$ is the linear parameter, $\theta_N = (b_1, b_2, \dots, b_k)^T$ is the nonlinear parameter and $\varphi_j(\theta_N; t_i)$ is a quadratic differentiable nonlinear function, which depends on θ_N and t_i . With the least squares method, it reduces to the following nonlinear function.

$$v(\theta_L, \theta_N) = \sum_{i=1}^m (y_i - \eta(\theta_L, \theta_N; t_i))^2 \quad (2)$$

which can be rewritten in the matrix form

$$v(\theta_L, \theta_N) = \|y - \Phi(\theta_N)\theta_L\|^2 \quad (3)$$

where $y = (y_1, y_2, \dots, y_m)^T$, $t = (t_1, t_2, \dots, t_m)^T$ and $\varphi_j(\theta_N; t) = (\varphi_j(\theta_N; t_1), \dots, \varphi_j(\theta_N; t_m))^T$. Additionally, $\Phi(\theta_N) = (\phi_1(\theta_N; t), \dots, \phi_n(\theta_N; t))$ is a matrix function, and $\|\cdot\|^2$ denotes the Euclidean norm.

The minimization problem (3) is a nonlinear least squares problem that does not consider the properties of the parameters. The Gauss–Newton (GN) algorithm [1], the

Levenberg–Marquardt (LM) algorithm [2,3] and the iterative embedding points algorithm [4] are commonly used to solve such problems. In fact, there are two sets of mutually dependent parameters in this model: the linear parameter and the nonlinear parameter. Golub and Pereyra [5,6] refer to this type of data-fitting problem as a separable nonlinear least squares (SNLLS) problem. In consideration of the specialized separable structure of this type of model, they proposed the variable projection (VP) algorithm. The general aim of implementing this algorithm is to eliminate linear parameters and obtain a simplified problem with only nonlinear parameters. For any given nonlinear parameter θ_N , θ_L can be obtained by solving the following linear least squares problem.

$$\theta_L = \underset{\theta_L}{\operatorname{argmin}} \|y - \Phi(\theta_N)\theta_L\|^2 \quad (4)$$

The least squares solution of Equation (4) is

$$\hat{\theta}_L = \left(\Phi(\theta_N)^T \Phi(\theta_N)\right)^{-1} \Phi(\theta_N)^T y = \Phi(\theta_N)^+ y \quad (5)$$

where $\Phi(\theta_N)^+$ is the pseudo-inverse of $\Phi(\theta_N)$. By substituting $\hat{\theta}_L$ into (3), we obtain

$$v(\theta_N) = \|y - \Phi(\theta_N)\Phi(\theta_N)^+ y\|^2 = \|(I - \Phi(\theta_N)\Phi(\theta_N)^+)y\|^2 = \|(I - P_{\Phi(\theta_N)})y\|^2 = \|P_{\Phi(\theta_N)}^\perp y\|^2 \quad (6)$$

where $P_{\Phi(\theta_N)} = \Phi(\theta_N)\Phi(\theta_N)^+$ is the orthogonal projection on the linear space spanned by the column vectors of $\Phi(\theta_N)$, and $P_{\Phi(\theta_N)}^\perp = I - P_{\Phi(\theta_N)}$ is the projector on the orthogonal complement of the column space of $\Phi(\theta_N)$.

Equation (6) is the revised residual function, in which the linear parameter is eliminated. The VP algorithm reduces the dimensionality of the parameter space and increases the probability of finding the global optimal solution. It is an effective method for solving the SNLLS problem.

The VP algorithm has been significantly improved and widely applied since it was proposed. For example, Kaufman [7] proposed an improved VP algorithm based on the trapezoidal decomposition of a matrix, and provided a simplified Jacobian matrix of the VP method. Ruhe et al. [8] analyzed the asymptotic convergence of the simplified VP algorithm. O’Leary and Rust [9] subsequently discovered that a VP algorithm with a complete Jacobian matrix requires fewer iterations than Kaufman’s simplified algorithm. Alternatively, Ruano et al. [10] proposed a more simplified Jacobian matrix for the VP algorithm, and an improved VP algorithm that entailed applying QR decomposition to the sparse nonlinear function matrix. They found that their method effectively increased computational efficiency. Gan et al. [11] compared the separation algorithms for different Jacobian matrices and concluded that a VP algorithm with the complete Jacobian matrix is more stable than the simplified algorithm proposed by Kaufman. Gan and Li [12] proposed a VP algorithm that utilizes the classic Gram–Schmidt (GS) matrix decomposition method to treat cases in which the number of observations is significantly larger than the number of linear parameters; their algorithm was found to reduce the computational cost. Alternatively, Chen et al. [13] employed a modified GS method in their development of a more robust VP algorithm for SNLLS problems; they reported that the combination of a VP algorithm and the L–M method is more effective than a combination of the VP algorithm and the G–N method.

Although many considerable efforts have been dedicated to developing methods to solve SNLLS problems, there are few studies on ill-conditioned problems with iterative processes. It can be seen that the least squares solution given as Equation (5) is conditioned on the fact that $\left(\Phi(\theta_N)^T \Phi(\theta_N)\right)^{-1} \Phi(\theta_N)^T$ is computable—that is, the matrix $\Phi(\theta_N)^T \Phi(\theta_N)$ is invertible. However, when the equation is ill-conditioned, the smallest characteristic root of the matrix $\Phi(\theta_N)^T \Phi(\theta_N)$ is near zero; therefore, the elements of $\left(\Phi(\theta_N)^T \Phi(\theta_N)\right)^{-1}$ will become significantly large. Small changes in the observation value will significantly affect

the least squares solution, causing it to become unstable. Thus, a regularization method needs to be introduced to convert the ill-posed problem into a relatively mild or benign problem [14–16] before proceeding. There are many regularization methods, including Tikhonov regularization (TR) [17,18], truncated singular value decomposition (TSVD) regularization [19,20], kernel-based regularization [21,22] and l_1 -regularization [23]. Generally, a regularization method is applied to improve the condition of the ill-conditioned matrix by introducing regularization parameters. Common methods for determining regularization parameters include ridge-tracing, generalized cross-check and L-curve-based methods. The TSVD method is significantly popular because it is relatively well developed, and can be applied to solve computationally complex problems. Zeng et al. [24] estimated the linear parameters in a separable least squares parameter optimization process by implementing regularization parameter detection techniques in the TR and TSVD methods. In their attempt to regularize SNLLS ill-posed problems using a VP algorithm, Chen et al. [25] developed a weighted generalized cross-validation method to determine TR parameters. The effectiveness of the algorithm was verified through experimentation. Wang et al. [26] separated linear and nonlinear regularization parameters using a singular value decomposition (SVD)-based VP algorithm. They also estimated the linear parameters by using both the LS method and an L-curve-based spectral correction method. Their experiments confirmed that the algorithm can effectively solve ill-conditioned problems.

In this paper, an improved TR optimization method is proposed for the parameter estimation problem of SNLLS models. With the VP method as the basic framework, the algorithm was developed to take into account the specialized structure of the linear combination of nonlinear functions and separate the linear and nonlinear parameters. The linear parameters are estimated by using an improved TR method, whereas the nonlinear parameters are optimized by using the LM method. Regarding the iterative process, the simplified Jacobian matrix proposed by Kaufman is implemented in the VP algorithm. Numerical simulations were performed to compare the improved TR method to the original TR and TSVD regularization methods. The effectiveness of the VP algorithm with the improved TR method was verified, and the accuracies of different linear parameter estimation methods were evaluated.

We begin this paper by summarizing the VP method and the existing problems related to solve SNLLS problems. The methods of nonlinear parameter estimation and linear parameter estimation are explained based on an improved VP algorithm derived from SVD, and the steps to solve SNLLS problems are then detailed. Thereafter, we compare and evaluate the performances of the method proposed in this paper and the conventional regularized VP algorithm by numerical simulations.

2. Parameter Estimation Method

2.1. Regularization Algorithm for Linear Parameter Estimation

With no loss of generality, the model is abbreviated as

$$y = \Phi\theta_L \quad (7)$$

Then, as a result of applying SVD, Φ can be rewritten as

$$\Phi = USV^T = \sum_{i=1}^l u_i \sigma_i v_i^T \quad (8)$$

where $U = (u_1, u_2, \dots, u_m)$ and $V = (v_1, v_2, \dots, v_n)$ are composed of the eigenvectors $\Phi^T\Phi$ and $\Phi\Phi^T$ respectively; and $U \in R^{m \times m}$, $V \in R^{n \times n}$, $U^T U = I$ and $V^T V = I$. $S = \text{diag}(\sigma_1, \dots, \sigma_l, \sigma_{l+1}, \dots, \sigma_n)$ is a diagonal matrix with diagonal elements that are singular values of Φ , where $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_l > 0$ and $\sigma_{l+1} = \sigma_{l+2} = \dots = \sigma_n = 0$. l is the rank of Φ .

The Moore–Penrose inverse of Φ is given as $\Phi^+ = \sum_{i=1}^l \frac{u_i^T v_i}{\sigma_i}$. The least squares solution of the linear parameter, according to Equation (5), is given as

$$\hat{\theta}_L = \sum_{i=1}^l \frac{u_i^T y v_i}{\sigma_i} \tag{9}$$

Clearly, if some small singular values are significantly close to zero, even a small error will produce a significantly large discrepancy between the least square solution and the true value, making the solution unstable. To avoid this problem, a filter factor was introduced to suppress the error term in the ill-conditioned solution. Thus, the approximate solution can be obtained as

$$\hat{\theta}_L = \sum_{i=1}^l \varphi \frac{u_i^T y v_i}{\sigma_i} \tag{10}$$

2.1.1. TSVD Method

The TSVD method removes the disturbing influence of small singular values on the LS solution, thereby allowing a stable solution to be achieved. This is achieved by setting a threshold and then setting any singular value smaller than this threshold to equal zero. Generally, an appropriate threshold can be determined by setting a cutoff ratio coefficient α , such that, if $\sigma_i < \alpha\sigma_1$, then σ_i is set to zero. The corresponding TSVD filter factor is

$$\varphi_\alpha = \begin{cases} 1 & \sigma_i \geq \alpha\sigma_1 \\ 0 & \sigma_i < \alpha\sigma_1 \end{cases} \tag{11}$$

At this point, the TSVD regularization solution is given as

$$\hat{\theta}_L = \sum_{i=1}^l \varphi_\alpha \frac{u_i^T y v_i}{\sigma_i} = \sum_{i=1}^k \frac{u_i^T y v_i}{\sigma_i} \tag{12}$$

where $k \leq l$ is the cutoff point for the singular value, which is typically implemented as the regularization parameter in the TSVD method.

2.1.2. TR Method

The TR method is currently the preferred method for solving ill-conditioned problems. The method entails the use of a regularization parameter μ to constrain the solution and construct a secondary optimization problem as follows:

$$\theta_L = \operatorname{argmin}_{\theta_L} \left\{ \|y - \Phi(\theta_N)\theta_L\|^2 + \frac{1}{2}\mu^2\|\theta_L\|^2 \right\} \tag{13}$$

where μ is the regularization parameter. When θ_N is fixed, the solution of Equation (13) becomes

$$\hat{\theta}_L = \left(\Phi(\theta_N)^T \Phi(\theta_N) + \mu^2 I \right)^{-1} \Phi(\theta_N)^T y \tag{14}$$

Subsequently, applying SVD to the matrix yields

$$\hat{\theta}_L = \sum_{i=1}^l \varphi_\mu \frac{u_i^T y v_i}{\sigma_i} = \sum_{i=1}^l \frac{\sigma_i^2}{\sigma_i^2 + \mu^2} \frac{u_i^T y v_i}{\sigma_i} \tag{15}$$

Thus, the TR filter factor can be expressed as $\varphi_\mu = \frac{\sigma_i^2}{\sigma_i^2 + \mu^2}$.

2.1.3. Improved TR Method

The TSVD and TR methods are essentially the same. However, there is a difference in the extent to which either method can reduce the influence of small singular values on the solution. Specifically, the TR method changes all singular values, which may cause the approximated solution to be over-smoothed. Alternatively, the TSVD method sets a small portion of the singular values to zero, which not only reduces the variance, but also affects the resolution of the solution, resulting in low accuracy. However, the improved TR method only changes small singular values by enabling the determination of the regularization matrix [27]. Given that the differences between the larger and smaller singular values of the ill-conditioned matrix are large, the singular values are determined to be small singular values with a significant influence when the sum of the standard components of small singular values accounts for more than 95% of the standard deviation. Consequently, they are regularized to reduce the influence on the standard deviation. This condition can be expressed as follows [28]:

$$\sum_{i=k}^l \frac{\sigma_0}{\sigma_i} \geq 95\% \sum_{i=1}^l \frac{\sigma_0}{\sigma_i} \tag{16}$$

Thus, the regularization parameter k can be obtained. Then, with the filter factor set as $\varphi_{k,\mu} = \begin{cases} 1 & 1 < i \leq k \\ \frac{\sigma_i^2}{\sigma_i^2 + \mu^2} & k < i \leq l \end{cases}$, when $1 \leq i \leq k$, $\varphi_{k,\mu} = \varphi_\alpha$; when $k \leq i \leq l$, $\varphi_{k,\mu} = \varphi_\mu$. The solution of the improved TR method is

$$\hat{\theta}_L = \sum_{i=1}^l \varphi_{k,\mu} \frac{u_i^T y v_i}{\sigma_i} \tag{17}$$

The improved TR method is a combination of the TSVD and TR methods. Unlike the TSVD method, the improved TR method does not disregard small singular values; furthermore, it yields a more accurate solution. Additionally, unlike the conventional TR method, the improved TR method only modifies the small singular values, keeping the large singular values unchanged, to ensure high accuracy.

A regularization method is employed to determine the appropriate regularization parameters. The L-curve method was adopted in this work. This method entails selecting different μ values to obtain a set of points $(\|y - \Phi \hat{\theta}_L\|^2, \|\hat{\theta}_L\|)$; subsequently, an L-shaped curve is constructed with $\|y - \Phi \hat{\theta}_L\|^2$ as the abscissa and $\|\hat{\theta}_L\|$ as the ordinate. Finally, the corresponding μ regularization parameter value is determined by locating the point of maximum curvature.

2.2. LM Algorithm for Nonlinear Parameter Estimation

After values for the linear parameters have been determined, Equation (6) contains only nonlinear parameters, the values of which can be determined by using the LM algorithm. The iteration strategy is

$$\theta_N^{k+1} = \theta_N^k + \alpha_k d_k \tag{18}$$

where α_k is the step length, which ensures that the objective function is in a descending state, and d_k is the search direction. α_k can be obtained by employing a line search method for imprecise searches. Let m_k be the smallest non-negative integer m satisfying $r(\theta_N^k + \rho^m d_k) \leq r(\theta_N^k) + \sigma \rho^m g_k^T d_k$ in the k^{th} iteration process. Then,

$$\alpha_k = \rho^{m_k} \tag{19}$$

where d_k can be determined by solving the following equation:

$$\left[J(\theta_N^k)^T J(\theta_N^k) + \gamma_k I \right] d_k = -J(\theta_N^k)^T F(\theta_N^k) \tag{20}$$

where γ_k is the damping parameter that controls the magnitude and direction of d_k . When γ_k is sufficiently large, the direction of d_k is consistent with the negative gradient direction of the objective function. When γ_k tends toward zero, d_k tends toward the G–N direction. In the LM algorithm implemented here, γ is adjusted by employing a strategy similar to adjusting the radius of the trust region [29]. A quadratic function is defined at the current iteration point as follows:

$$q(d) = r(\theta_N^k) + (J(\theta_N^k)F(\theta_N^k))^T d + \frac{1}{2}d^T (J(\theta_N^k)F(\theta_N^k))d \tag{21}$$

The ratio of the objective function to the increment of $q(d)$ is denoted by η_k as follows:

$$\eta_k = \frac{\Delta r(d_k)}{\Delta q(d_k)} = \frac{r(\theta_N^{k+1}) - r(\theta_N^k)}{(J(\theta_N^k)F(\theta_N^k))^T d_k + \frac{1}{2}d_k^T (J(\theta_N^k)F(\theta_N^k))d_k} \tag{22}$$

In each step of the LM algorithm, γ_k is first assigned an initial value to calculate d_k , such as the corresponding previous iteration value. Then, γ_k is adjusted according to the value of η_k ; d_k is subsequently calculated according to the adjusted γ_k , and a line search is performed. Clearly, when η_k is close to one, γ should be relatively smaller; this can be achieved by using the LM method to solve the nonlinear problem. When η_k is close to zero, the modulus length of d_k must be reduced, and γ should be relatively larger. When η_k is neither close to one nor zero, the value of the parameter γ_k is determined to be suitable. The critical values of η are typically 0.25 and 0.75. Accordingly, the update rule of γ_k is given as

$$\gamma_{k+1} := \begin{cases} 0.1\gamma_k, & \eta_k > 0.75, \\ \gamma_k, & 0.25 \leq \eta_k \leq 0.75, \\ 10\gamma_k, & \eta_k < 0.25. \end{cases} \tag{23}$$

In Equation (20), $J(\theta_k)$ is the Jacobian matrix of the residual function, which, according to the simplified Jacobian matrix calculation method proposed by Kaufman [7], is given as

$$J_{KAU} = -P_\Phi^\perp D\Phi\Phi^- y \tag{24}$$

where $D\Phi$ is the Fréchet derivative of the matrix Φ , and Φ^- is the symmetric generalized inverse of Φ . In [30], $P_\Phi = \Phi\Phi^-$, where Φ^- satisfies:

$$\Phi\Phi^-\Phi = \Phi, (\Phi\Phi^-)^T = \Phi\Phi^- \tag{25}$$

Thus, Φ^+ is not needed to calculate P_Φ .

Applying SVD to decompose matrix Φ yields

$$\Phi = USV^T = [U_1, U_2] \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} [V_1, V_2]^T \tag{26}$$

where Σ is the l th diagonal matrix, and the diagonal elements are the singular values of Φ . Note that $\text{rank}(U_1) = \text{rank}(V_1) = l$, where l is the rank of Φ . Accordingly, we obtain

$$\Phi^- = [V_1, V_2] \begin{bmatrix} \Sigma^{-1} & 0 \\ 0 & 0 \end{bmatrix} [U_1, U_2]^T = V_1\Sigma^{-1}U_1^T \tag{27}$$

Then

$$P_\Phi = \Phi\Phi^- = [U_1, U_2] \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} [U_1, U_2]^T \tag{28}$$

Thus, the corresponding residual function is

$$r(\theta_N) = \|P_\Phi^\perp y\|^2 = \|[U_1, U_2] \begin{bmatrix} 0 & 0 \\ 0 & I_{m-r} \end{bmatrix} [U_1, U_2]^T y\|^2 = \|U_2 U_2^T y\|^2 \tag{29}$$

and the Jacobian matrix is

$$J_{KAU} = -P_{\Phi}^{\perp} D\Phi\Phi^{-1}y = -U_2U_2^T D\Phi V_1\Sigma^{-1}U_1^T y \tag{30}$$

2.3. Algorithm Solution Determination

Here, we describe an improved TR optimization method for the SNLLS problem. This method entails the implementation of the VP algorithm to separate the variables, followed by the use of the improved TR method to update the linear parameters, and the LM method to search for nonlinear parameters. We compared the performance of the improved TR regularization method to those of the conventional TR method and TSVD regularization method to verify the effectiveness of the method. The model is given as

$$(\hat{\theta}_L, \hat{\theta}_N) = \arg \left\{ \min_{\theta_N} \|y - \Phi(\theta_N)\tilde{\theta}_L\|^2, \tilde{\theta}_L = \sum_{i=1}^l \varphi \frac{u_i^T y v_i}{\sigma_i} \right\} \tag{31}$$

A summary of the steps used for obtaining the solution is given as follows:

- Step 1: Take the initial value of the nonlinear parameter $\theta_N^{(0)}$, the maximum number of iteration steps M and set $k = 0$.
- Step 2: The initial nonlinear parameter value $\theta_N^{(0)}$ is used to calculate the initial values of the linear parameters $\theta_L^{(0)}$ via the TR, TSVD or improved TR method. Then the residual function $r(\theta_N)$ and approximate Jacobian matrix J_{KAU} are obtained.
- Step 3: The iterative step length α_k and search direction \mathbf{d}_k are determined by solving Equations (19) and (20), respectively; thereafter, the nonlinear parameters are updated according to Equation (18).
- Step 4: The linear and nonlinear parameters are cross-updated until $k > M$; then, the calculation is terminated.

3. Numerical Simulation

3.1. Predicting the Mackey-Class Time Series Using an RBF Neural Network

Numerical Mackey–Glass time-series simulations were performed for the validation experiment. In the experiment, the LM algorithm was separately implemented in the TSVD-based VP method (VP_{TSVD}), the TR-based VP method (VP_{TR}) and the improved TR method ($VP_{TSVD-TR}$) to fit a time-series image. This experiment was also performed to reveal the advantages and disadvantages of the aforementioned VP algorithms. The experimental environment was MATLAB R2016a running on a 1.80 GHz PC with Windows 7.

The exponential model is given as

$$\eta(a, \alpha; x) = a_1 + \sum_{j=2}^n a_j e^{-\lambda_j \|x - z_j\|^2} \tag{32}$$

where $a = (a_1, a_2, \dots, a_n)$ is the linear parameter set, and $\alpha = (\lambda_2, \lambda_3, \dots, \lambda_n; z_2^T, z_3^T, \dots, z_n^T)^T$ is the nonlinear parameter set, for the functions $\varphi_1(\alpha; x) \equiv 1$, $\varphi_j(\alpha; x) = e^{-\lambda_j \|x - z_j\|^2}$, and $j = 2, 3, \dots, n$. This model was used to fit the chaotic Mackey–Glass time series generated by the following delay differential equation:

$$\frac{dy(t)}{dt} = \frac{ay(t - \tau)}{1 + y^c(t - \tau)} - by(t) \tag{33}$$

where $a = 0.2, b = 0.1, c = 10, \tau = 17$. The initial value condition was set as $y(0) = 1.2$, and the time interval was set to 0.1. The Runge–Kutta method was used to solve the differential equation and generate 500 data points, as shown in Figure 1. Among these 500 data points, 300 were extracted from the generated data as follows:

$$[y(t - 24), y(t - 18), y(t - 12), y(t - 6), y(t)] (t = [24 \quad 323])$$

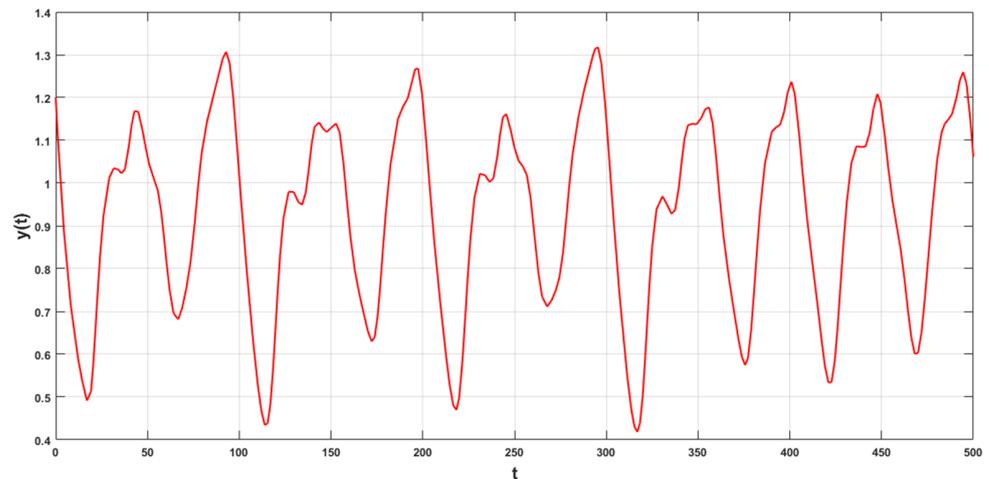


Figure 1. Chaotic Mackey–Glass time series.

Using these 300 data points, $VP_{TSVD+LM}$, VP_{TR+LM} and $VP_{TSVD-TR+LM}$ were applied to estimate the parameters for the model; the remaining data were used for prediction.

When $n = 2$, the exponential model yielded 24 nonlinear parameters and 19 linear parameters. Given the same initial iterative value, the fits of the curves derived from the training and prediction data points output by the three algorithms are shown in Figure 2. The red circles correspond to data points extracted from the time-series images. It can be intuitively determined from the figure that the curve fitted by the $VP_{TSVD-TR+LM}$ algorithm is in good agreement with the data generated.

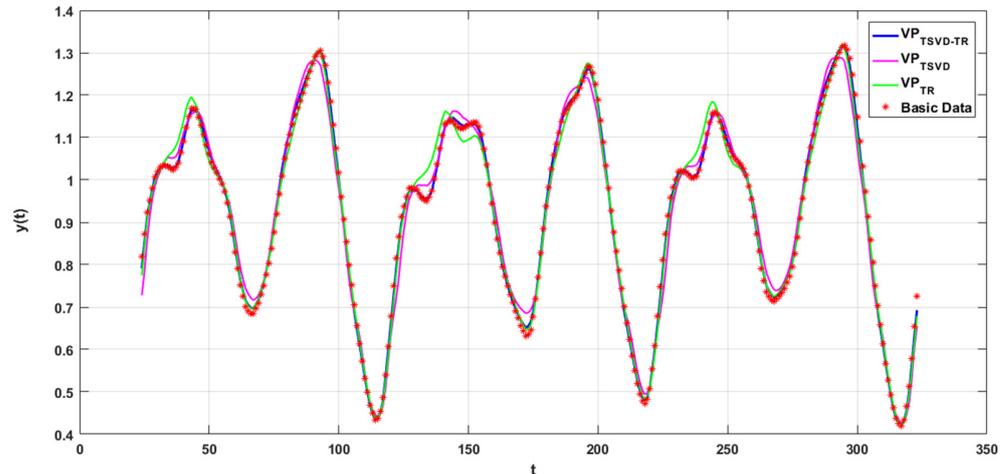


Figure 2. Curve-fitting evaluation results.

Table 1 presents the results—the number of iterations, the number of calculated functions, the second norm of the vector formed by the residuals of each point and the root mean square error (RMSE t) of the objective function with respect to the original data—for the three iterative methods. Figure 3 shows the convergences of the corresponding nonlinear parameters for the three iterative algorithms.

From the results in Table 1, we can see that the VP_{TSVD} method required the most iterations, and had the highest RMSE- t value. However, the VP_{TR} and the $VP_{TSVD-TR}$ methods were improved significantly and optimized in the iterative process due to fewer iterations, fewer function calculations and smaller RMSEs. Although the numbers of iterations and function evaluations were similar, the results obtained by using the $VP_{TSVD-TR}$ method to calculate the model parameters were considerably more accurate.

Table 1. Simulated results for a Mackey–Glass time series.

	Number of Iterations	Function Evaluation Number	Second Norm of Residual Vector	RMSE-t
VP_{TSVD}	75	1953	0.34373	0.27918
VP_{TR}	5	135	0.36427	0.19919
$VP_{TSVD-TR}$	7	185	0.05818	0.09653

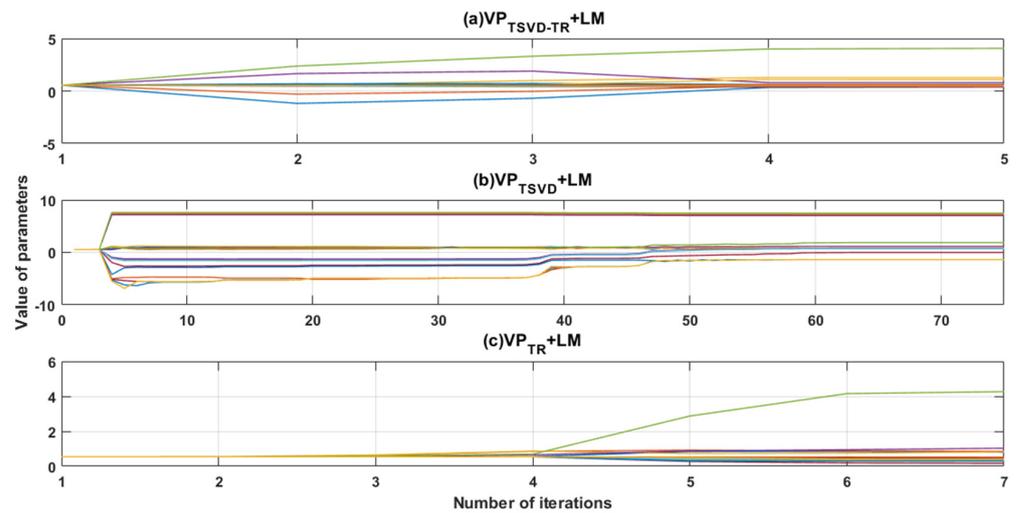


Figure 3. The iterative process for nonlinear parameters.

From Figure 3, it is evident that all three methods yielded converging results with the same initial values. However, the results of the $VP_{TSVD}+LM$ method began to approach the optimized solution around the 40th iteration, whereas the $VP_{TSVD-TR}+LM$ algorithm began slowly approaching the optimized parameter values at the third iteration, indicating a higher rate of convergence.

Every 10 points were grouped, beginning at the 400th point in the Mackey–Glass time series, to obtain several time-series predictions. The RMSE results for each group are shown in Figure 4.

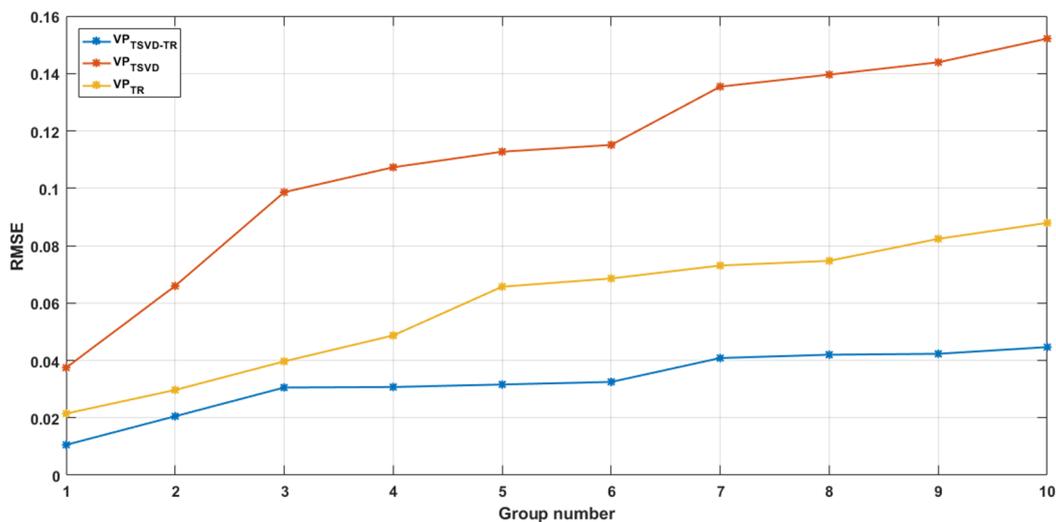


Figure 4. RMSE results following multi-step time-series forecasting.

It can be seen from Figure 4 that the RMSE values for the $VP_{TSVD}+LM$ method increased with the number of predictions, indicating decreasing prediction accuracy and

degrading prediction ability. Conversely, in the case of the VP_{TSVD-TR}+LM method, there was a minimal increase in the RMSE, proving the superior stability of the method. These results also indicate that the proposed method features a higher prediction accuracy.

3.2. Height Anomaly Fitting

A height anomaly is the difference in elevation between the surface of the reference ellipsoid and a quasi-geoid. As we can see in Figure 5, the height anomaly ζ can be expressed as $\zeta = H - H_{normal}$. The geodetic height of a point can be calculated by the GPS positioning technique. The normal height can be calculated as long as the height anomaly is determined in the certain area.

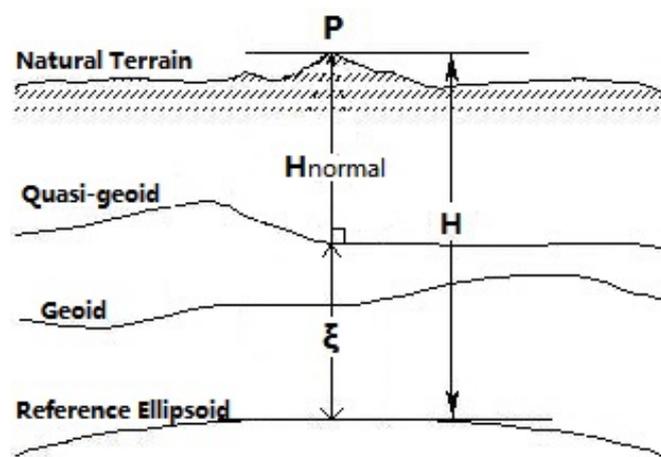


Figure 5. A geodetic height, normal height and height anomaly.

A multi-surface function model was used, and its expression had the characteristic that the linear and nonlinear parameters could be separated. The TSVD method, TR method and improved algorithm were applied to solve the parameters of the model according to the height anomaly values of known control points. The model is best expressed as:

$$\zeta_i = \sum_{i=1}^n a_i Q_i(X, Y, X_i, Y_i) \tag{34}$$

where $Q_i(X, Y, X_i, Y_i) = \sqrt{(X - X_i)^2 + (Y - Y_i)^2 + b_i^2}$. The example data, coordinates, geodetic heights, normal heights and height anomaly values of the known points are shown in Table 2.

D01–D15 were used for fitting, and the rest were used to verify the reliability of the parameters and the prediction abilities of the three algorithms. The iteration initial values of the nonlinear parameters were set to 0.5. Model (34) and the algorithm proposed in this paper were applied to fit the height anomalies, and the results are compared with the results calculated by VP_{TSVD} and VP_{TR}.

The parameter estimations were obtained after alternative calculations between the linear parameters and nonlinear parameters. The fitting images corresponding to the three algorithms are shown in Figure 6. Table 3 shows the residual sum of squares, root mean squared error and coefficient of determination values during the processes of fitting and prediction. It can be concluded from Table 3 that all algorithms enabled the model to get high fitting accuracy. During fitting and prediction, The order of magnitude of the root mean square error for VP_{TSVD}, VP_{TR} and VP_{TSVD-TR} is respectively 10^{-1} , 10^{-2} and 10^{-3} . The accuracy of VP_{TSVD} algorithm was slightly lower. The fitting accuracies of VP_{TR} and VP_{TSVD-TR} were higher. It can be seen from SSR_f and $RMSE_f$ that the improved algorithm proposed in this paper had better performance in height anomaly prediction.

Table 2. Coordinates, geodetic heights (H), normal heights (H_{normal}) and height anomaly values (ξ).

No.	Latitude	Longitude	H (m)	H_{normal} (m)	ξ (m)
D01	343000.0000	1120000.0000	1001.5220	1056.5490	-55.0270
D02	343000.0000	1120230.0000	1009.1290	1063.9280	-54.7990
D03	343000.0000	1120500.0000	1012.3950	1067.2510	-54.8560
D04	343000.0000	1120730.0000	1070.1860	1125.3030	-55.1170
D05	343000.0000	1121000.0000	1025.2330	1080.2320	-54.9990
D06	343000.0000	1121230.0000	1019.4310	1074.4540	-55.0230
D07	343000.0000	1121500.0000	1026.7090	1081.7560	-55.0470
D08	343000.0000	1121730.0000	1067.9940	1123.1420	-55.1480
D09	343000.0000	1122000.0000	1157.7220	1212.5890	-54.8670
D10	343000.0000	1122230.0000	1017.5320	1072.6350	-55.1030
D11	343000.0000	1122500.0000	1004.0630	1058.9510	-54.8880
D12	343000.0000	1122730.0000	1004.3500	1059.1140	-54.7640
D13	342730.0000	1120000.0000	988.8130	1043.3570	-54.5440
D14	342730.0000	1120230.0000	975.3010	1029.7370	-54.4360
D15	342730.0000	1120500.0000	983.5130	1038.1040	-54.5910
D16	342730.0000	1120730.0000	988.8550	1043.5930	-54.7380
D17	342730.0000	1121000.0000	1108.9430	1163.7680	-54.8250
D18	342730.0000	1121230.0000	980.5940	1035.2260	-54.6320
D19	342730.0000	1121500.0000	964.1870	1018.5260	-54.3390
D20	342730.0000	1120000.0000	1001.5220	1056.5490	-55.0270

Table 3. Residual sum of squares (SSR, SSR_f), root mean squared error (RMSE, $RMSE_f$) and coefficient of determination (R^2) values.

Algorithms	SSR (m ²)	SSR_f (m ²)	RMSE (m)	$RMSE_f$ (m)	R^2
VP _{TSVD}	0.5004	0.9337	0.1826	0.4321	0.2599
VP _{TR}	0.0850	0.3770	0.0753	0.2746	0.8744
VP _{TSVD-TR}	0.0798	0.3545	0.0729	0.2663	0.8820

The normal height of each point was obtained according to $H_{normal} = H - \xi$ and the height anomaly values fitted by three variable projection algorithms. The parameter estimations were verified by the known normal height. Figure 7 shows the fitting effects, and Table 4 lists the normal height fitting residual and RMSE of each point. It can be concluded from Figure 7 and Table 4 that the differences between the normal heights estimated by the three algorithms and the known example data are very small, and the three broken lines are almost coincident, which indicates that the model well fit the normal heights of known points. This is consistent with the conclusions drawn in Figure 6 and Table 3.

Table 4. Fitting residuals (r) and RMSEs of the normal heights.

No.	H_{normal} (m)	r_{TSVD} (m)	r_{TR} (m)	$r_{TSVD+TR}$ (m)
D01	1056.5490	0.2871	0.1041	0.0902
D02	1063.9280	-0.0457	-0.0457	-0.0161
D03	1067.2510	0.1498	-0.0245	-0.0742
D04	1125.3030	0.1702	0.0015	0.0224
D05	1080.2320	-0.2271	-0.0266	0.0108
D06	1074.4540	-0.1912	-0.0084	-0.0399
D07	1081.7560	-0.0025	-0.0036	-0.0052
D08	1123.1420	0.2312	0.0942	0.1011
D09	1212.5890	0.0406	-0.1688	-0.1694
D10	1072.6350	0.2850	0.1092	0.1078
D11	1058.9510	-0.0049	-0.0108	-0.0069
D12	1059.1140	-0.2656	-0.0069	-0.0091
D13	1043.3570	-0.1411	-0.0576	-0.0701
D14	1029.7370	-0.2104	-0.0699	-0.0269
D15	1038.1040	-0.0726	0.1139	0.0854
D16	1043.5930	-0.1204	0.1882	0.0878
D17	1163.7680	-0.2611	0.2026	0.1069
D18	1035.2260	-0.4721	-0.0523	-0.1058
D19	1018.5260	-0.6528	-0.4140	-0.4319
D20	1056.5490	-0.4494	-0.3555	-0.3710
RMSE	0	0.2678	0.1520	0.1474

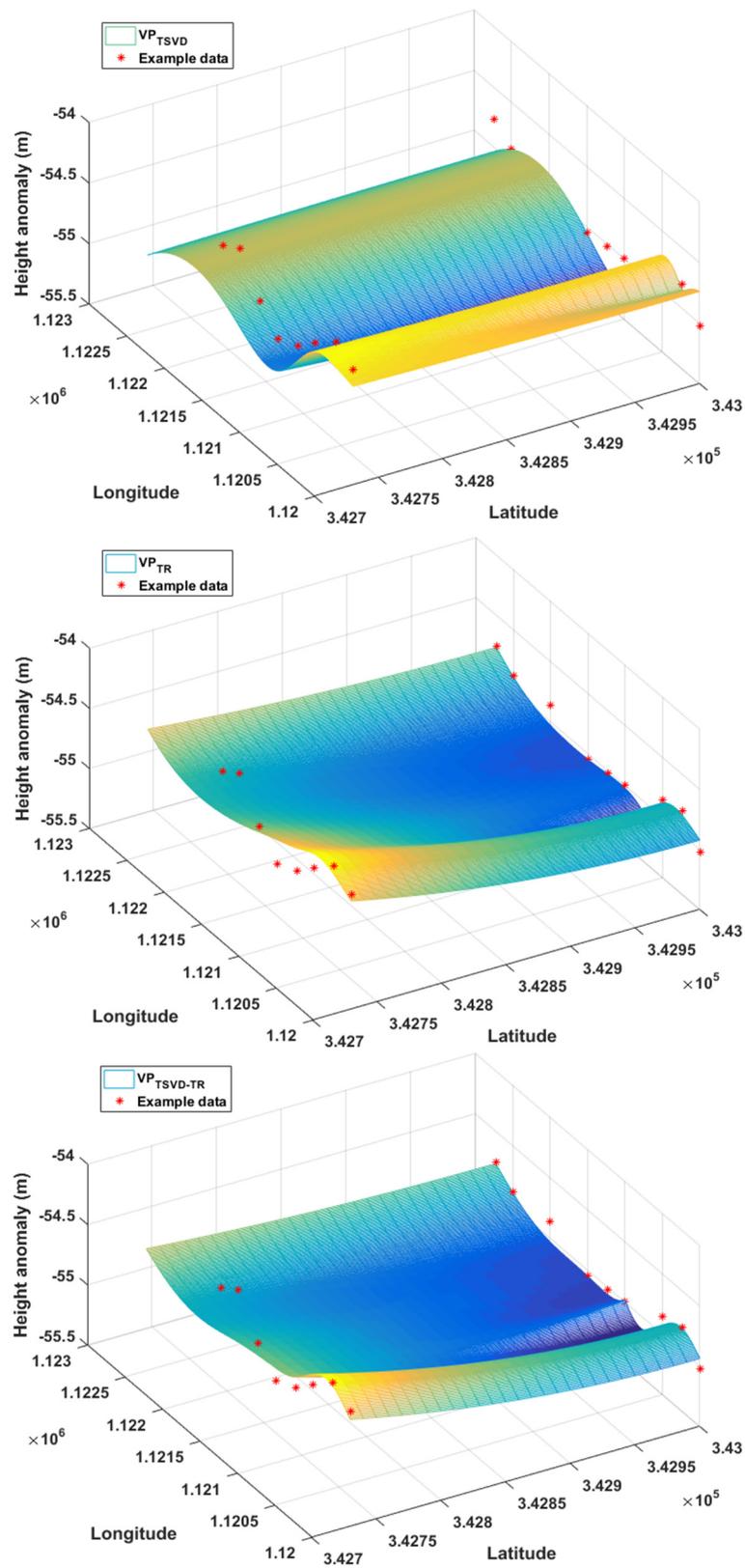


Figure 6. Fitting effects with respect to height anomaly values.

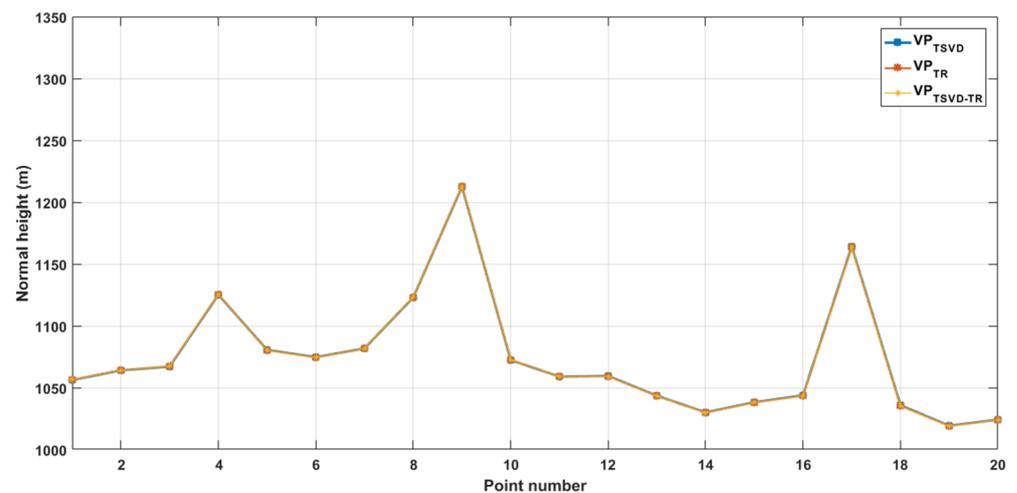


Figure 7. Normal height fitting effects.

4. Conclusions

Separable nonlinear models have been widely applied in many fields and studied by many scholars all over the world. The VP method is one of the most effective algorithms for solving related problems. However, the classical VP method may not be able to estimate the parameters of ill-conditioned problems. As regularization is the most common method for solving ill-conditioned problems, we proposed an improved regularized VP method for the SNLLS problem. This method entails applying certain rules to determine the filter factors necessary to correct small singular values, and using the L-curve method to determine regularization parameters. The results of numerical simulations revealed that the mean square error of the proposed regularized VP method was less than those of the TSVD and TR methods. Furthermore, we found that combining a regularization method with the LM algorithm-based nonlinear parameter estimation method results in a higher rate of convergence. This improved regularized parameter estimation method was demonstrated to have certain advantages in terms of its ability to mitigate the ill-conditioned problem and improve parameter estimation accuracy. It was also found to be more stable when applied to an SNLLS problem.

Author Contributions: Methodology, H.G.; software, L.W.; writing—original draft, H.G.; writing—review and editing, G.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (Grant no. 42074009).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: 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 is no conflict of interest regarding the publication of this paper.

References

- Okatani, T.; Deguchi, K. On the Wiberg algorithm for matrix factorization in the presence of missing components. *Int. J. Comput. Vis.* **2007**, *72*, 329–337. [\[CrossRef\]](#)
- Levenberg, K. A method for the solution of certain non-linear problems in least squares. *Q. Appl. Math.* **1944**, *2*, 164–168. [\[CrossRef\]](#)
- Marquardt, D.W. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *SIAM J. Appl. Math.* **1963**, *11*, 431–441. [\[CrossRef\]](#)

4. Hong, J.H.; Zach, C.; Fitzgibbon, A. Revisiting the variable projection method for separable nonlinear least squares problems. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 5939–5947.
5. Golub, G.H.; Pereyra, V. The Differentiation of Pseudo-Inverses and Nonlinear Least Squares Problems Whose Variables Separate. *SIAM J. Numer. Anal.* **1973**, *10*, 413–432. [[CrossRef](#)]
6. Golub, G.H.; Pereyra, V. Separable nonlinear least squares: The variable projection method and its applications. *Speech Commun.* **2003**, *45*, 63–87. [[CrossRef](#)]
7. Kaufman, L. A variable projection method for solving separable nonlinear least squares problems. *BIT Numer. Math.* **1975**, *15*, 49–57. [[CrossRef](#)]
8. Ruhe, A. Algorithms for separable nonlinear least squares problems. *SIAM Rev.* **1980**, *22*, 318–337. [[CrossRef](#)]
9. O’Leary, D.P.; Rust, B.W. Variable projection for nonlinear least squares problems. *Comput. Optim. Appl.* **2013**, *54*, 579–593. [[CrossRef](#)]
10. Ruano, A.E.B.; Jones, D.I.; Fleming, P.J. A new formulation of the learning problem of a neural network controller. In Proceedings of the 30th IEEE Conference on Decision and Control, Brighton, UK, 11–13 December 1991; pp. 865–866.
11. Gan, M.; Chen, C.L.P.; Chen, G.; Chen, L. On some separated algorithms for separable nonlinear least squares problems. *IEEE Trans. Cybern.* **2018**, *48*, 2866–2874. [[CrossRef](#)]
12. Gan, M.; Li, H. An Efficient Variable Projection Formulation for Separable Nonlinear Least Squares Problems. *IEEE Trans. Cybern.* **2014**, *44*, 707–711. [[CrossRef](#)]
13. Chen, G.Y.; Gan, M.; Ding, F. Modified Gram-Schmidt Method Based Variable Projection Algorithm for Separable Nonlinear models. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *30*, 2410–2418. [[CrossRef](#)]
14. Böckmann, C. A modification of the trust-region Gauss-Newton method to solve separable nonlinear least squares problems. *J. Math. Syst. Estim. Control* **1995**, *5*, 1–16.
15. Chung, J.J.; Nagy, G. An efficient iterative approach for large-scale separable nonlinear inverse problems. *SIAM J. Sci. Comput.* **2010**, *31*, 4654–4674. [[CrossRef](#)]
16. Li, X.L.; Liu, K.; Dong, Y.S.; Tao, D.C. Patch alignment manifold matting. *IEEE Trans. Neural Netw. Learn. Syst.* **2018**, *29*, 3214–3226. [[CrossRef](#)]
17. Tikhonov, A.N.; Arsenin, V.Y. Solutions of ill-posed problems. *SIAM Rev.* **1979**, *21*, 266–267.
18. Park, Y.; Reichel, L.; Rodriguez, G. Parameter determination for Tikhonov regularization problems in general form. *J. Comput. Appl. Math.* **2018**, *343*, 12–25. [[CrossRef](#)]
19. Hansen, P.C. The Truncated SVD as A Method for Regularization. *BIT Numer. Math.* **1987**, *27*, 534–553. [[CrossRef](#)]
20. Xu, P.L. Truncated SVD methods for Discrete Linear Ill-posed Problems. *Geophys. J. Int.* **1998**, *1335*, 505–514. [[CrossRef](#)]
21. Aravkin, A.Y.; Drusvyatskiy, D.; van Leeuwen, T. Efficient quadratic penalization through the partial minimization technique. *IEEE Trans. Autom. Control* **2018**, *63*, 2131–2138. [[CrossRef](#)]
22. Pang, S.C.; Li, T.; Dai, F.; Yu, M. Particle Swarm Optimization Algorithm for Multi-Salesman Problem with Time and Capacity Constraints. *Appl. Math. Inf. Sci.* **2013**, *7*, 2439–2444. [[CrossRef](#)]
23. Wang, X.Z.; Liu, D.Y.; Zhang, Q.Y.; Huang, H.L. The iteration by correcting characteristic value and its application in surveying data processing. *J. Heilongjiang Inst. Technol.* **2001**, *15*, 3–6.
24. Zeng, X.Y.; Peng, H.; Zhou, F.; Xi, Y.H. Implementation of regularization for separable nonlinear least squares problems. *Appl. Soft Comput.* **2017**, *60*, 397–406. [[CrossRef](#)]
25. Chen, G.Y.; Gan, M.; Chen CL, P.; Li, H.X. A Regularized Variable Projection Algorithm for Separable Nonlinear Least-Squares Problems. *IEEE Trans. Autom. Control* **2019**, *64*, 526–537. [[CrossRef](#)]
26. Wang, K.; Liu, G.L.; Tao, Q.X.; Zhai, M. Efficient Parameters Estimation Method for the Separable Nonlinear Least Squares Problem. *Complexity* **2020**, *2020*, 9619427. [[CrossRef](#)]
27. Fuhry, M.; Reichel, L. A new Tikhonov regularization method. *Number Algorithms* **2012**, *59*, 433–445. [[CrossRef](#)]
28. Lin, D.f.; Zhu, J.j.; Song, Y.C. Regularized singular value decomposition parameter construction method. *J. Surv. Mapp.* **2016**, *45*, 883–889.
29. Byrd, R.H.; Schnabel, R.B.; Shultz, G.A. A Trust Region Algorithm for Nonlinearly Constrained Optimization. *SIAM J. Numer. Anal.* **1987**, *24*, 1152–1170. [[CrossRef](#)]
30. Golub, G.H.; Styan, G.P.H. Numerical computations for univariate linear models. *J. Stat. Comput. Simul.* **1973**, *2*, 253–274. [[CrossRef](#)]