Iterative parameter estimation of nonlinear systems

Miao Yin, Li Zhao*
School of Automation and Electronic Engineering, Qingdao University of Science and Technology, Qingdao, 266061, China
Information Section, Qingdao Hospital of Traditional Chinese Medicine, Qingdao, 266033, China
E-mail: (Corresponding author) ljwan@qust.edu.cn

Abstract: This paper discusses the identification methods for nonlinear systems. Using the negative gradient search, the gradient based iterative algorithm is derived to determine the parameters of the nonlinear system. In addition, in order to improve the computational efficiency, the gradient-based iterative algorithm based on the model transformation is derived. The basic principle is to transform a complex nonlinear system into a linear or simple nonlinear system. Finally, a numerical example is given to verify the effectiveness of the proposed methods.

1. Introduction
System identification deals with the problem of building mathematical models of systems based on observation data [1, 2]. It has widespread applications in many areas, such as biological, chemical production process, bioreactor, internal combustion engine, etc. System identification has become an important branch in the field of control science [3].

Iterative identification algorithm and recursive identification algorithm are two important and basic parameter estimation methods [4–6]. The identification of iterative algorithm is usually related to the number of iterations, independent of the time, so it is usually used to identify the number of systems offline. The basic idea of iterative identification is as follows: based on data with fixed length, interactive estimation theory and hierarchical identification theory are applied to refresh parameter estimation. Since there are unknown terms, the information vector can be calculated with the parameter estimation value of the previous iteration, and then the calculated information vector can be used for parameter estimation [7, 8].

In real life, the model structure of the system is usually more complex and the identification of nonlinear systems has been widely concerned [9–13]. For example, Narendra K et al. proposed an iterative method for the identification of nonlinear systems from samples of inputs and outputs in presence of noise [14]. Ding et al. presented the identification method of Hammerstein nonlinear ARMAX systems [15, 16]. Recently, Noor et al. proposed iterative methods for solving nonlinear equations by using the homotopy perturbation approach [17] and Wu et al. presented an iterative algorithm for solving complex conjugate and transpose matrix equations [18].

Gradient iteration is one of the common iterative methods for solving nonlinear equations. Li et al. proposed an iterative algorithm based on the gradient method, which uses the negative gradient search method to determine the parameters of nonlinear systems [19]. On the basis of the above works, this paper derives a gradient-based iterative estimation algorithm based on the negative gradient search. This algorithm results in a large amount of computation. Therefore, in order to reduce the amount of
computation, this paper gives the gradient-based iterative algorithm based on the model transformation. The basic idea is to transform a complex nonlinear system into a simpler one.

The rest of this paper is arranged as follows. Section 2 introduces the identification model to be discussed in this paper. In Section 3, the gradient-based iterative algorithm for nonlinear systems is derived. In Section 4, an iterative method for nonlinear systems based on the model transformation are proposed. A numerical example is provided in Section 5 to illustrate the proposed methods. Finally, Section 6 makes a summary.

2. The system description
Consider a nonlinear system which is described as the following expression,

\[ y = \frac{bx + c}{x + a}, \]  

(1)

Where \(\{x_i, y_i, i = 1, 2, \ldots, n\} \quad (n \geq 3)\) are the measurement data polluted by noise, \(a, b\) and \(c\) are unknown parameters to be estimated. This paper drives \(a, b\) and \(c\) from the data set by minimizing the following cost functions,

\[ J_1(a, b, c) = \frac{1}{2} \sum_{i=1}^{n} \left( y_i - \frac{bx_i + c}{x_i + a} \right)^2. \]  

(2)

In order to minimize \(J_1(a, b, c)\), the general way is to let its partial derivatives with respect to \(a, b\) and \(c\) be equal to 0, that is

\[ \frac{\partial J_1(a, b, c)}{\partial a} = \sum_{i=1}^{n} \left( y_i - \frac{bx_i + c}{x_i + a} \right) \frac{bx_i + c}{(x_i + a)^2} = 0, \]

\[ \frac{\partial J_1(a, b, c)}{\partial b} = -\sum_{i=1}^{n} \left( y_i - \frac{bx_i + c}{x_i + a} \right) \frac{x_i}{x_i + a} = 0, \]

\[ \frac{\partial J_1(a, b, c)}{\partial c} = -\sum_{i=1}^{n} \left( y_i - \frac{bx_i + c}{x_i + a} \right) \frac{1}{x_i + a} = 0. \]

From the above nonlinear equations, it is very difficult to obtain the values of \(a, b\) and \(c\) by solving the equations directly. Therefore, in order to avoid solving the nonlinear equations directly, we use the iterative identification method to estimate the parameter values \(a, b\) and \(c\) at the extreme point of \(J_1(a, b, c)\).

3. The gradient iterative algorithm
Because of its low computational complexity, the gradient method has attracted much attention in solving large-scale nonlinear optimization problems [20, 21]. In particular, in applications such as image processing, data analysis and machine learning, people need approximate solutions with certain precision, therefore, various gradient algorithms have been proposed [22–27]. In addition, the gradient method is useful for identification [28]. Many gradient-based parameter identification methods have been developed using the auxiliary model, the multi-innovation theory [29] and the data filtering [30, 31].

The iterative identification method is an important branch of system identification, which can be realized by using the gradient search, the least squares principle, and the Newton optimization. The gradient iterative algorithm is a traditional optimization algorithm to find the minimum value of a cost function. Compared with the least-squared-based iterative algorithm, the gradient-based iterative algorithm has less computational burden, so it has been widely used in system identification [32–35].
In this section, a gradient-based iterative estimation algorithm is derived and the iterative solutions of \( a, b, \) and \( c \) are obtained using the negative gradient search.

The parameter vector is defined,

\[
\mathbf{g} := \begin{bmatrix} a \\ b \\ c \end{bmatrix}
\]  

(3)

The gradient of \( J_1(\mathbf{g}) := J_1(a,b,c) \) in regard to \( \mathbf{g} \) is

\[
\nabla J_1(\mathbf{g}) = \begin{bmatrix} \frac{\partial J_1(a,b,c)}{\partial a} \\ \frac{\partial J_1(a,b,c)}{\partial b} \\ \frac{\partial J_1(a,b,c)}{\partial c} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} \left( y_i - \frac{bx_i + c}{x_i + a} \right) \frac{bx_i + c}{(x_i + a)^2} \\ -\sum_{i=1}^{n} \left( y_i - \frac{bx_i + c}{x_i + a} \right) x_i \\ -\sum_{i=1}^{n} \left( y_i - \frac{bx_i + c}{x_i + a} \right) x_i + a \end{bmatrix} .
\]

Let \( k = 1,2,3,\cdots \) be an iteration variable, \( \hat{\mathbf{g}} := \begin{bmatrix} \hat{a}_k \\ \hat{b}_k \\ \hat{c}_k \end{bmatrix} \) denote the estimate of \( \mathbf{g} \) at iteration \( k \). For the optimization problem in (2), the gradient-based iterative algorithm is presented by minimizing \( J_1(\mathbf{g}) \) as follows.

\[
\hat{\mathbf{g}}_{k+1} = \hat{\mathbf{g}}_k + \mu_k \mathbf{q}_k ,
\]

(4)

\[
\mu_k = \arg\min_{\mu \geq 0} J_1\left( \hat{\mathbf{g}}_k + \mu \mathbf{q}_k \right).
\]

(5)

\[
\mathbf{q}_k = -\nabla J_1(\hat{\mathbf{g}}_k) .
\]

(6)

\[
\nabla J_1(\hat{\mathbf{g}}_k) = \begin{bmatrix} \sum_{i=1}^{n} \left( y_i - \frac{\hat{b}_k x_i + \hat{c}_k}{x_i + \hat{a}_k} \right) \frac{\hat{b}_k x_i + \hat{c}_k}{(x_i + \hat{a}_k)^2} \\ -\sum_{i=1}^{n} \left( y_i - \frac{\hat{b}_k x_i + \hat{c}_k}{x_i + \hat{a}_k} \right) x_i \\ -\sum_{i=1}^{n} \left( y_i - \frac{\hat{b}_k x_i + \hat{c}_k}{x_i + \hat{a}_k} \right) x_i + \hat{a}_k \end{bmatrix} .
\]

(7)

According to the above formulas, the calculation steps of the gradient-based iterative algorithm are summarized as follows:

1. Collect the measured data \( (x_i, y_i), i = 1,2,\cdots,n \).
2. To initialize, let \( k = 1, \hat{a}_1, \hat{b}_1\) and \( \hat{c}_1 \) be arbitrary real numbers, and the preset a small \( \varepsilon \).
3. Compute \( \nabla J_1(\hat{\mathbf{g}}_k) \) by (7).
4. Compute the searching direction \( \mathbf{q}_k \) by (6), determine the step-size \( \mu_k \) by (5).
5. Compute \( \hat{\mathbf{g}}_{k+1} \) by (4). If \( \| \hat{\mathbf{g}}_{k+1} - \hat{\mathbf{g}}_k \| > \varepsilon \), increase \( k \) by 1 and go to step 3; otherwise, obtain the estimate \( \hat{\mathbf{g}}_{k+1} \) and terminate the procedure.
4. Iterative methods based on the model transformation

The system (1) is transformed and solved in this paper. The details are as follows. By making a transformation, system (1) is rewritten as the following form,

\[(x + a)y = bx + c, \quad (8)\]

and define a new cost function,

\[J_2(\vartheta) = \frac{1}{2} \sum_{i=1}^{n} (x_i y_i + ay_i - bx_i - c)^2. \quad (9)\]

So its partial derivatives are

\[
\frac{\partial J_2(\vartheta)}{\partial a} = \sum_{i=1}^{n} y_i (x_i y_i + ay_i - bx_i - c) = 0, \\
\frac{\partial J_2(\vartheta)}{\partial b} = -\sum_{i=1}^{n} x_i (x_i y_i + ay_i - bx_i - c) = 0, \\
\frac{\partial J_2(\vartheta)}{\partial c} = -\sum_{i=1}^{n} (x_i y_i + ay_i - bx_i - c) = 0.
\]

The gradient of \(J_2(\vartheta)\) in regard to \(\vartheta\) is

\[
\nabla J_2(\vartheta) = \left[ \begin{array}{c} \frac{\partial J_2(\vartheta)}{\partial a} \\ \frac{\partial J_2(\vartheta)}{\partial b} \\ \frac{\partial J_2(\vartheta)}{\partial c} \end{array} \right] = \left[ \begin{array}{c} \sum_{i=1}^{n} y_i (x_i y_i + ay_i - bx_i - c) \\ -\sum_{i=1}^{n} x_i (x_i y_i + ay_i - bx_i - c) \\ -\sum_{i=1}^{n} (x_i y_i + ay_i - bx_i - c) \end{array} \right].
\]

So, we can obtain the gradient-based iterative algorithm based on the model transformation as follows:

\[
\vartheta_{k+1} = \vartheta_{k} + \mu_k d_k, \quad (10)
\]

\[
\mu_k = \arg \min_{\mu \geq 0} J_1(\vartheta_{k} + \mu d_k), \quad (11)
\]

\[
d_k = -\nabla J_1(\vartheta_{k}), \quad (12)
\]

\[
\nabla J_2(\vartheta) = \left[ \begin{array}{c} \sum_{i=1}^{n} y_i (x_i y_i + ay_i - bx_i - c) \\ -\sum_{i=1}^{n} x_i (x_i y_i + ay_i - bx_i - c) \\ -\sum_{i=1}^{n} (x_i y_i + ay_i - bx_i - c) \end{array} \right]. \quad (13)
\]

From the above analysis and calculation, it can be seen that the gradient iteration method based on model transformation is simpler than the gradient iteration method, and the calculation amount is reduced.

5. Example

This section discusses the following nonlinear system,

\[y = \frac{bx + c}{x + a} + v, \quad (14)\]
where \( \nu \) is the measurement errors (i.e., noise). The measurement data are \( \{ x_i, y_i, i = 1, 2, \ldots, n \} \). In simulation, we take \( a = 2.2, b = 0.8 \) and \( c = 2.3 \), the input \( x_i \) as a signal sequence, \( \nu \) as a white noise sequence with zero mean and variance \( \sigma^2 \).

Applying the proposed iterative algorithms to estimate the parameters of this example system with different noise variances, the parameter estimates and their errors are shown in Tables 1-4, and the parameter estimation errors versus \( k \) are shown in Figs. 1-2.

### Table I. The Gradient Estimates and Errors with \( \sigma^2 = 0.5^2 \)

| \( k \) | \( a \)  | \( b \)  | \( c \)  | \( \delta(\%) \) |
|-------|--------|--------|--------|----------------|
| 0     | 1.85000| 1.45000| 2.60000| 24.28170       |
| 500   | 2.18441| 1.12245| 2.35585| 9.98325        |
| 1000  | 2.20102| 0.95533| 2.33737| 4.86817        |
| 1500  | 2.20672| 0.87596| 2.32529| 2.44818        |
| 2000  | 2.20878| 0.83759| 2.31845| 1.30376        |
| 3000  | 2.20940| 0.80984| 2.31246| 0.56234        |
| 5000  | 2.20771| 0.80185| 2.30869| 0.35839        |
| 8000  | 2.20448| 0.80179| 2.30515| 0.21508        |
| True values | 2.20000| 0.80000| 2.30000|                |

### Table II. The Gradient Estimates and Errors with \( \sigma^2 = 1.0^2 \)

| \( k \) | \( a \)  | \( b \)  | \( c \)  | \( \delta(\%) \) |
|-------|--------|--------|--------|----------------|
| 0     | 1.85000| 1.45000| 2.60000| 24.28170       |
| 500   | 2.18441| 1.12245| 2.35585| 10.93750       |
| 1000  | 2.20102| 0.95533| 2.33737| 6.25347        |
| 1500  | 2.20672| 0.87596| 2.32529| 4.03630        |
| 2000  | 2.20878| 0.83759| 2.31845| 2.96863        |
| 3000  | 2.20940| 0.80984| 2.31246| 2.19307        |
| 5000  | 2.20771| 0.80185| 2.30869| 1.94202        |
| 8000  | 2.20448| 0.80179| 2.30515| 1.89976        |
| True values | 2.20000| 0.80000| 2.30000|                |

Fig. 1: The Gradient-based iterative estimation errors \( \delta \) versus \( k \) with \( \sigma^2 = 0.5^2 \) and \( \sigma^2 = 1.0^2 \).

Table 1, Table 2 and Figure 1 reflect the variation of the parameter estimation error and iteration times of the gradient-based iterative algorithm under different variances. It can be seen from the curve variation diagram, with the increase of iteration, the error become smaller and smaller. This indicates...
that the parameter values obtained by the iterative algorithms are closer to the real values, and the smaller the variance is, the closer to the real values the parameter values are.

**TABLE III.** THE GRADIENT-BASED ITERATIVE ALGORITHM BASED ON THE MODEL TRANSFORMATION WITH $\sigma^2 = 0.5^2$

| k  | a         | b         | c         | $\delta$ (%) |
|----|-----------|-----------|-----------|--------------|
| 0  | 1.85000   | 1.45000   | 2.60000   | 40.69384     |
| 500| 2.18441   | 1.12245   | 2.35585   | 15.18831     |
| 1000| 2.20102  | 0.95533   | 2.33737   | 8.46534      |
| 1500| 2.20672  | 0.87596   | 2.32529   | 4.77861      |
| 2000| 2.20878  | 0.83759   | 2.31845   | 2.71561      |
| 3000| 2.20940  | 0.80984   | 2.31246   | 0.89356      |
| 5000| 2.20771  | 0.80185   | 2.30869   | 0.11729      |
| 8000| 2.20448  | 0.80179   | 2.30515   | 0.07877      |
| True values | 2.20000 | 0.80000   | 2.30000   |              |

**TABLE IV.** THE GRADIENT-BASED ITERATIVE ALGORITHM BASED ON THE MODEL TRANSFORMATION WITH $\sigma^2 = 1.0^2$

| k  | a         | b         | c         | $\delta$ (%) |
|----|-----------|-----------|-----------|--------------|
| 0  | 1.85000   | 1.45000   | 2.60000   | 40.69384     |
| 500| 2.18441   | 1.12245   | 2.35585   | 15.90173     |
| 1000| 2.20102  | 0.95533   | 2.33737   | 9.53646      |
| 1500| 2.20672  | 0.87596   | 2.32529   | 6.05397      |
| 2000| 2.20878  | 0.83759   | 2.31845   | 4.10949      |
| 3000| 2.20940  | 0.80984   | 2.31246   | 2.39625      |
| 5000| 2.20771  | 0.80185   | 2.30869   | 1.67095      |
| 8000| 2.20448  | 0.80179   | 2.30515   | 1.58914      |
| True values | 2.20000 | 0.80000   | 2.30000   |              |

Fig. 2: The Model transformation gradient iterative estimation errors $\delta$ versus $k$ with $\sigma^2 = 0.5^2$ and $\sigma^2 = 1.0^2$

Table 3 and table 4 reflect the parameter estimation error of the gradient-based iterative algorithm based on the model transformation under different variances. It can be seen from the results that the algorithm is more efficient than the gradient iterative algorithm.
6. Conclusions
In this paper, based on the gradient search, the gradient-based iterative algorithm is studied. In addition, in order to simplify the computation of gradient-based iterative algorithm, the gradient iterative algorithm based on the model transformation is studied. Finally, the simulation results show that the proposed algorithms are effective, and they can be extended to other nonlinear systems.

Acknowledgments
I would like to express my gratitude to all those who have helped me in the process of writing this paper.

First of all, I would like to thank my tutor Mr. Yang Shuguo for his help. He has been encouraging and guiding me to complete all stages of this paper. Without his consistent guidance and inspiration, this paper could not have reached the present form.

Secondly, I would also like to pay tribute to my teacher Wan Lijuan, who has put forward helpful suggestions and suggestions for my paper. I am very grateful to her for her help in completing this paper.

Finally, I would like to thank my dear family for their care and trust in me over the years. I also want to thank my friends and classmates who gave me help and time during the difficult process of my paper and helped me solve my problems.

References
[1] F. Ding, System identification--Iterative search principle and identification method [M], Science Press, Beijing 2018. 8.
[2] Isermann R, Műnchhof M, Identification of dynamic systems: An identification with applications [M], Springer Science Business Media 2010.
[3] F. Ding, System Identification-New Theory and Methods [M], Science Press, Beijing 2013. 1.
[4] L. J. Wan, F. Ding, Decomposition-and gradient-based iterative identification algorithms for multivariable systems using the multi-innovation theory, Circuits Systems and Signal Processing 38 (7) (2019) 2971-2991.
[5] Y. Ji, X. K. Jiang, L. J. Wan, Hierarchical least squares parameter estimation algorithm for two-input Hammerstein finite impulse response systems, Journal of the Franklin Institute 357 (8) (2020) 5019-5032.
[6] Y. Ji, C. Zhang, Z. Kang, T. Yu, Parameter estimation for block-oriented nonlinear systems using the key term separation, International Journal of Robust and Nonlinear Control 30 (9) (2020) 3727-3752.
[7] L. Xu, F. Ding, Recursive least squares and multi-innovation stochastic gradient parameter estimation methods for signal modeling, Circuits Systems and Signal Processing 36 (4) (2017) 1735-1753.
[8] M. H. Li, X. M. Liu, Maximum likelihood least squares based iterative estimation for a class of bilinear systems using the data filtering technique, International Journal of Control Automation and Systems 18 (6) (2020) 1581-1592.
[9] H. Nijmeijer, On the theory of nonlinear control systems [M], Three Decades of Mathematical System Theory 1989.
[10] H. Habbi, M. Kidouche, M. Zelmat, Data-driven fuzzy models for nonlinear identification of a complex heat exchanger, Applied Mathematical Modelling 35 (3) (2011) 1470-1482.
[11] F. Ding, P.X. Liu, G. Liu, Identification methods for Hammerstein nonlinear systems, Digital Signal Processing 21 (2) (2011) 215-238.
[12] J. Vörös, Parameter identification of Wiener systems with multisegment piecewise-linear nonlinearities. Systems & Control Letters 56 (2) (2007) 99-105.
[13] F. Ding, Y. Shi, T. Chen, Auxiliary model-based least-squares identification methods for Hammerstein output-error systems, Systems & Control Letters 56 (5) (2007) 373-380.
[14] K. Narendra, P. Gallman, An iterative method for the identification of nonlinear systems using a Hammerstein model, IEEE Transactions on Automatic Control 11 (3) (2003) 546-550.
[15] F. Ding, T. Chen, Identification of Hammerstein nonlinear ARMAX systems, Automatica 41 (9) (2005) 1479-1489.
[16] F. Ding, Y. Shi, T. Chen, Gradient-based identification methods for Hammerstein nonlinear ARMAX models, Nonlinear Dynamic 45 (1-2) (2006) 31-43.
[17] M.A. Noor, W.A. Khan, New iterative methods for solving nonlinear equation by using homotopy perturbation method, Applied Mathematics Computation 219 (8) (2012) 3565-3584.
[18] A.G. Wu, L.L. Lv, G.R. Duan, Iterative algorithms for solving a class of complex conjugate and transpose matrix equations, Applied Mathematics Computation 217 (21) (2011) 8343-8353.
[19] J.H. Li, R.F. Ding, Y. Yang, Iterative parameter identification methods for nonlinear functions, Applied Mathematical Modelling 36 (2012) 2739-2750.
[20] S. Ghadimi, G. Lan, Accelerated gradient methods for non-convex nonlinear and stochastic programming, Mathematical Programming 156 (1-2) (2015) 59-99.
[21] G. Lan, An optimal method for stochastic composite optimization, Mathematical Programming 133 (2012) 365-397.
[22] L.Z. Liao, L. Qi, H.W. Tam, A gradient-based continuous method for large-scale optimization problems, Journal of Global Optimization 31 (2) (2005) 271-286.
[23] Y. Nesterov, Universal gradient methods for convex optimization problems, Mathematical Programming 152 (2015) 381-404.
[24] Y. Nesterov, Smooth minimization of non-smooth functions, Mathematical Programming 103 (1) (2005) 127-152.
[25] R. Tavakoli, H. Zhang, A nonmonotone spectral projected gradient method for large-scale topology optimization, Numerical Algebra Control & Optimization 2 (2012) 395-412.
[26] A. Beck, M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, SIAM Journal on Imaging Sciences 2 (1) (2009) 183-202.
[27] Y. Chen, W.W. Hager, M. Yashtini, X. Ye, H. Zhang, Bregman operator splitting with variable step size for total variation image reconstruction, Computational Optimization and Applications 54 (2) (2013) 317-342.
[28] F. Ding, L. Xu, Q. M. Zhu, Performance analysis of the generalised projection identification for time-varying systems, IET Control Theory and Applications 10 (18) (2016) 2506-2514.
[29] L. Xu, F. Ding, The parameter estimation algorithms for dynamical response signals based on the multi-innovation theory and the hierarchical principle, IET Signal Processing 11 (2) (2017) 228-237.
[30] Y.J. Wang, F. Ding, Novel data filtering based parameter identification for multiple-input multiple-output systems using the auxiliary model, Automatica 71 (2016) 308-313.
[31] Y.J. Wang, F. Ding, The filtering based iterative identification for multivariable systems, IET Control Theory and Applications 10 (8) (2016) 894-902.
[32] J.S. Li, Y.Y. Zheng, Z.P. Lin, Recursive identification of time varying systems: Self-tuning and matrix RLS algorithms, Systems & Control Letters 66 (2014) 104-110.
[33] B.Q. Mu, H.F. Chen, L.Y. Wang, G. Yin, W.X. Zheng, Recursive identification of hammerstein systems: convergence rate and asymptotic normality, IEEE Transactions on Automatic Control 62 (7) (2017) 3277-3292.
[34] A.G. Wu, W.X. Zhang, Y. Zhang, An iterative algorithm for discrete periodic Lyapunov matrix equations, Automatica 87 (2018) 395-403.
[35] D.H. Zhao, Y.S. Wei, Y.T. Liu, Spectrum optimization via FFT-based conjugate gradient method for unimodular sequence design, Signal Processing 142 (2018) 354-365.