Time-varying Forgetting Factor Stochastic Gradient Algorithm for Nonlinear Systems with Color Noise

Chunming Xu
School of Mathematics and Statistical, Yancheng Teachers University, 224051, Yancheng, PR China

Abstract This paper considers the identification of nonlinear systems with color noise, and introduces a new time-varying forgetting factor based stochastic gradient (TVFF-SG) algorithm to estimate the system parameters. The basic idea of the time-varying forgetting factor is that when the algorithm starts, we give the forgetting factor a relative smaller value, which will speed up the convergence. In addition, the forgetting factor will increase slowly as time goes on so that the convergence procedure of the model will be more stable. Simulation results are presented to demonstrate the effectiveness of the proposed algorithm.

Keywords System Identification, Stochastic Gradient Algorithm, Time-varying Forgetting Factor, Nonlinear Systems, Color Noise

1 Introduction

System identification aims to build a mathematical model from data generated by a dynamical system. System identification is a powerful technique which has great potential in many applied areas such as model-based simulation, prediction and control of dynamical systems[1–5]. As a result, the study of system identification has attracted the attention of researchers from many scientific disciplines such as chemical processes [6], fault identification [7], biological process [8] and power system[9], etc.

In the past several decades, a variety of types of reliable system identification algorithms have been proposed such as bayesian method [10], stochastic gradient method [11], least squares method [12], Newton iterative method [13], evolutionary optimization based method [14] and so on.

Among the methods mentioned above, the stochastic gradient (SG) algorithm is widely used in many system identification and parameter estimation problems due to its simplify [15–17]. However, a major limitation of SG algorithm is that it has a slower convergence speed and lower convergence accuracy. In order to solve this problem, forgetting factor stochastic gradient(FF-SG) algorithm was proposed[18–20]. The FF-SG algorithm introduces a forgetting factor into SG so that the convergence rate will be increased. However, it is very difficult to choose a suitable forgetting factor. When considering the forgetting factor, an important idea is to make a tradeoff between the convergence rate and the stability of convergence. Based on this idea, the researchers seek to apply variable forgetting factors for solving the parameter identification problems. Till now, several different types of variable forgetting factors have been developed to improve the identification performance[21–23].

In practical engineering situations, many systems exhibit nonlinear phenomena, and the researchers aim to take advantage of nonlinear phenomena to solve practical problems. For example, the detection and estimation of parameter faults in nonlinear systems with nonlinear fault distribution functions is studied and a detection and estimation tool is proposed and applied to the one-wheel model with lumped friction[24]. The authors invented a new engineering low-frequency noise oscillators utilizing the nonlinear phenomena of micromechanical resonators[25]. The single degree of freedom electromagnetic energy is studied in [26] and it is find that we can make use of the duffing-type nonlinearities to reduce the size of energy harvesting devices while at the same time keeping their power output. Therefore, there is an immense need for algorithms that can effectively identify the parameters of the nonlinear systems for theoretical research and practical application. In this paper, a time-varying forgetting factor stochastic gradient algorithm is proposed to identify the parameters of nonlinear systems with color noise.

2 Problem Formulation

Consider the following nonlinear system with color noise:

\[ y(t) = A(z)f(y(t)) + B(z)u(t) + \frac{D(z)}{C(z)} v(t) \]  

where \( u(t) \) is the system input, \( v(t) \) is a random white noise with zero mean and variance \( \delta^2 \), \( y(t) \) is the measured system output, and \( f(t) \) is a nonlinear function, \( A(z), B(z), C(z) \)
and $D(z)$ are the polynomials in the unit backward shift operator $z^{-1}y(t) = y(t-1)$ with:

$$A(z) := a_1z^{-1} + a_2z^{-2} + ... + a_nz^{-n} \quad (2)$$

$$B(z) := b_1z^{-1} + b_2z^{-2} + ... + b_nz^{-n} \quad (3)$$

$$C(z) := 1 + c_1z^{-1} + c_2z^{-2} + ... + c_nz^{-n} \quad (4)$$

$$D(z) := 1 + d_1z^{-1} + d_2z^{-2} + ... + d_nz^{-n} \quad (5)$$

Assume that the order $n_a$, $n_b$, $n_c$ and $n_d$ are known and $u(t) = 0$, $v(t) = 0$ and $y(t) = 0$ as $t \leq 0$. Define the inner variable:

$$w(t) = \frac{D(z)}{C(z)}v(t) \quad (6)$$

which is an autoregressive moving average process. Define the parameter vectors $\theta$, $\theta_s$, $\theta_n$ and the information vectors $\psi(t), \psi_s(t), \psi_n(t)$ as:

$$\theta = \begin{bmatrix} \theta_s \\ \theta_n \end{bmatrix} \quad (7)$$

$$\theta_s = [a_1, a_2, ..., a_n, b_1, b_2, ..., b_n]^T \in \mathbb{R}^{n \times 1} \quad (8)$$

$$\theta_n = [c_1, c_2, ..., c_n, d_1, d_2, ..., d_n]^T \in \mathbb{R}^n \quad (9)$$

$$n_1 = n_a + n_b \quad (10)$$

$$n = n_a + n_b + n_c + n_d \quad (11)$$

$$\psi(t) = \begin{bmatrix} \psi_s(t) \\ \psi_n(t) \end{bmatrix} \quad (12)$$

$$\psi_s(t) = [-f(y(t-1)), -f(y(t-2)), ..., -f(y(t-n_a)), u(t-1), u(t-2), ..., u(t-n_b)]$$

$$\psi_n(t) = [-w(t-1), -w(t-2), ..., -w(t-n_c), v(t-1), v(t-2), ..., v(t-n_d)] \quad (13)$$

Then Eq.(6) can be written as:

$$w(t) = [(1-C(z))w(t)+D(z)v(t)] = \psi_n^T(t)\theta_n + v(t) \quad (15)$$

Based on Eq.(6) and Eqs.(7-14), Eq.(1) can be recast as follows:

$$y(t) = \psi_s^T(t)\theta_s + w(t) = \psi^T(t)\theta + v(t) \quad (16)$$

This is the identification model of the nonlinear system with color noise.

### 3 Time-varying Forgetting Factor Stochastic Gradient Algorithm

Define the cost functions:

$$J(\theta) = (y(t) - \psi^T(t)\theta)^2 \quad (17)$$

Let $\hat{\theta}(t)$ be the estimate of $\theta$ at time $t$. Using the gradient search and minimizing $J(\theta)$, we can obtain the following forgetting factor stochastic gradient algorithm for estimating the parameter vector $\hat{\theta}(t)$:

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \frac{\psi(t)}{r(t)}[y(t) - \psi^T(t)\hat{\theta}(t-1)] \quad (18)$$

$$r(t) = r(t-1) + \lambda(t) \|\psi(t)\|^2, r(0) = 1 \quad (19)$$

However, the information vector $\psi(t)$ contains the unknown intermediate variables $w(t-i), i = 1, 2, ..., n_a$ and $v(t-i), i = 1, 2, ..., n_d$, the algorithm in (18)-(19) can not be implemented. The solution is to replace the unknown variables by their corresponding estimates $\hat{w}(t-i), i = 1, 2, ..., n_a$ and $\hat{v}(t-i), i = 1, 2, ..., n_d$. Thus the forgetting factor stochastic gradient (TVFF-SG) algorithm can be described as follows:

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \frac{\hat{\psi}(t)}{r(t)}[y(t) - \hat{\psi}^T(t)\hat{\theta}(t-1)] \quad (20)$$

$$r(t) = r(t-1) + \lambda(t) \|\hat{\psi}(t)\|^2, r(0) = 1 \quad (21)$$

$$\hat{\theta}(t) = \begin{bmatrix} \hat{\theta}_s(t) \\ \hat{\theta}_n(t) \end{bmatrix} \quad (22)$$

$$\hat{\psi}(t) = \begin{bmatrix} \hat{\psi}_s(t) \\ \hat{\psi}_n(t) \end{bmatrix} \quad (23)$$

$$\hat{\psi}_s(t) = [-\hat{w}(y(t-1)), -f(y(t-2)), ..., -f(y(t-n_a)), u(t-1), u(t-2), ..., u(t-n_b)]$$

$$\hat{\psi}_n(t) = [-\hat{w}(t-1), -\hat{w}(t-2), ..., -\hat{w}(t-n_c), \hat{v}(t-1), \hat{v}(t-2), ..., \hat{v}(t-n_d)] \quad (24)$$

$$\hat{\psi}_s(t) = [\hat{w}(y(t-1)), f(y(t-2)), ..., f(y(t-n_a)), u(t-1), u(t-2), ..., u(t-n_b)]$$

$$\hat{\psi}_n(t) = [-\hat{w}(t-1), -\hat{w}(t-2), ..., -\hat{w}(t-n_c), \hat{v}(t-1), \hat{v}(t-2), ..., \hat{v}(t-n_d)] \quad (25)$$

$$\hat{\psi}_s(t) = [\hat{w}(y(t-1)), f(y(t-2)), ..., f(y(t-n_a)), u(t-1), u(t-2), ..., u(t-n_b)]$$

$$\hat{\psi}_n(t) = [-\hat{w}(t-1), -\hat{w}(t-2), ..., -\hat{w}(t-n_c), \hat{v}(t-1), \hat{v}(t-2), ..., \hat{v}(t-n_d)] \quad (26)$$

$$\hat{\theta}_s(t) = [\hat{a}_1(t), \hat{a}_2(t), ..., \hat{a}_{n_s}(t), \hat{b}_1, \hat{b}_2, ..., \hat{b}_{n_s}(t)]^T \quad (28)$$

$$\hat{\theta}_n(t) = [\hat{c}_1(t), \hat{c}_2(t), ..., \hat{c}_{n_n}(t), \hat{d}_1, \hat{d}_2, ..., \hat{d}_{n_n}(t)] \quad (29)$$
where $\lambda(t)$ is the forgetting factor. If $\lambda(t) = a$, $a$ is a fixed constant between 0 and 1, the algorithm is the forgetting factor stochastic gradient algorithm. When $\lambda(t) = 1$, the algorithm is reduced to the standard stochastic gradient algorithm.

Studies have shown that choosing a suitable forgetting factor is of great importance for system identification. If the forgetting factor is very close to 1, it should need rather long time to find the true parameters of the systems. However, when the forgetting factor is small, the stability of the algorithm will be reduced. In this paper, to solve this problem, we propose a new time-varying forgetting factor stochastic gradient algorithm to estimate the system parameters. Formally, the time-varying forgetting factor is written as:

$$
\lambda(t) = \begin{cases} 
a + b_L & \text{if } (a + b_L) < 1 \\
1 & \text{otherwise}
\end{cases} \tag{30}
$$

where $a$ and $b$ are two constants, $0 \leq a, b < 1$, $L$ is the iteration length of the algorithm. When $t$ is very small, the forgetting factor is near to $a$, which will speed up the convergence. As time goes on, the forgetting factor is very close or equal to 1, so that the flowing convergence procedure of the model will be more stable.

The steps involved in computing the parameter estimate $\hat{\theta}(t)$ as $t$ increases using the TVFF-SG algorithm is summarized as follows:

1. Let $t = 1$, set the initial values $\hat{\theta}(0) = 1_n/p_0$, where $p_0$ is a large number (e.g., $p_0 = 10^6$).
2. Collect the input data $u(t)$ and the output data $y(t)$.
3. Form $\psi_s(t)$ and $\psi_n(t)$ by (24) and (25), respectively. Form $\hat{\psi}(t)$ using Eq.(23).
4. Compute $\lambda(t)$ using Eq.(30).
5. Update the parameter estimate $\hat{\theta}(t)$ using Eq.(20).
6. Compute $\hat{\phi}(t)$ and $\tilde{\phi}(t)$ using (26) and (27), respectively.
7. Increase $t$ by 1 and goto step 3.

### 4 Example

This section provides an example to show the effectiveness of the proposed TVFF-SG algorithm for the nonlinear systems, compared with the SG, FF-SG algorithms. Consider the following nonlinear system with color noise:

$$
y(t) = A(z)f(y(t)) + B(z)u(t) + \frac{D(z)}{C(z)}v(t) \tag{31}
$$

$$
A(z) := -0.59z^{-1} + 0.45z^{-2} \tag{32}
$$

$$
B(z) := 0.9z^{-1} + 0.62z^{-2} \tag{33}
$$

$$
C(z) := 1 - 0.25z^{-1} + 0.27z^{-2} \tag{34}
$$

$$
D(z) := 1 + 0.3z^{-1} - 0.2z^{-2} \tag{35}
$$

Then the parameter vector $\theta$ can be expressed below:

$$
\theta = [-0.59, 0.45, 0.9, 0.62, -0.25, 0.27, 0.3, -0.2]^T \tag{36}
$$

In simulation, the inputs $u(t)$ is taken as an uncorrelated persistent excitation signal sequence with zero mean and unit variance, and $v(t)$ as a white noise sequence with zero mean and variance $\delta_v^2 = 0.2^2$. Applying the SG algorithm, FF-SG algorithm and TVFF-SG algorithm to estimate the parameters of this system. For FF-SG, the forgetting factor is selected as $FF = 0.99$. For TVFF-SG, the parameters are selected as $a = 0.99, b = 0.05$.

Numerical results are displayed in tables 1, 2, 3 and figure 1. The parameter estimates and their errors with different innovation length are shown in tables 1, 2 and 3, and the parameter estimation errors $\delta = ||\hat{\theta}(t) - \theta||/||\theta||$ vs. $t$ are shown in figure 1. In addition, figure 2 shows the time evolution curves of the system parameters vs. $t$ of the TVFF-SG method.

The experiments have been conducted to demonstrate the properties of the proposed approach. Some attractive points can be summarized as follows:

1. (1) The parameter estimation errors of the three algorithms become smaller and smaller as $t$ increasing.
2. (2) SG doesn’t consider the forgetting factor so that it gets the poorest results.
3. (3) FF-SG and TVFF-SG both use the forgetting factors to speed up the convergence so they achieve better results than SG.
4. (4) TVFF-SG estimates have higher accuracy than the other two techniques, etc. SG and FF-SG. This is because that it considers both the tracking ability and the stability of the algorithm. This also confirms that the proposed TVFF-SG is effective for parameter identification.

### 5 Conclusions

This paper investigates the parameter identification of the nonlinear system using a novel time-varying forgetting factor stochastic gradient(TVFF-SG) algorithm. The proposed TVFF-SG method considers both the tracking ability and the stability of the algorithm and takes a proper tradeoff between them. The proposed method is simple and easily to implement in system identification problems. The simulation results verify the effectiveness of the proposed approach. The presented method can also be used to solve other system identification problems.

In order to improve the performance of the algorithm, we can combine it with other system identification methods. For example, we can extend the proposed algorithm to multi-innovation based identification method[27] to improve the parameter estimation accuracy. The proposed method can also combine the filtering technology to identify the parameters of the other nonlinear systems.

### Acknowledgements

The authors are grateful to the anonymous reviewers for their valuable comments and suggestions that have helped to improve the presentation of the paper. This work is par-
Table 1. The SG estimates and errors.

<table>
<thead>
<tr>
<th>t</th>
<th>$a_1$</th>
<th>$a_2$</th>
<th>$b_1$</th>
<th>$b_2$</th>
<th>$c_1$</th>
<th>$c_2$</th>
<th>$d_1$</th>
<th>$d_2$</th>
<th>$\delta$(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>-0.25643</td>
<td>0.10755</td>
<td>0.41384</td>
<td>0.33316</td>
<td>-0.14659</td>
<td>0.07092</td>
<td>0.16205</td>
<td>-0.05569</td>
<td>56.30533</td>
</tr>
<tr>
<td>200</td>
<td>-0.26419</td>
<td>0.12450</td>
<td>0.44625</td>
<td>0.34670</td>
<td>-0.14576</td>
<td>0.08137</td>
<td>0.16283</td>
<td>-0.06424</td>
<td>53.54414</td>
</tr>
<tr>
<td>500</td>
<td>-0.28536</td>
<td>0.14612</td>
<td>0.48627</td>
<td>0.36103</td>
<td>-0.15559</td>
<td>0.09841</td>
<td>0.17536</td>
<td>-0.07677</td>
<td>49.49292</td>
</tr>
<tr>
<td>1000</td>
<td>-0.30569</td>
<td>0.17490</td>
<td>0.51704</td>
<td>0.36992</td>
<td>-0.14771</td>
<td>0.10897</td>
<td>0.17071</td>
<td>-0.08672</td>
<td>46.33699</td>
</tr>
<tr>
<td>2000</td>
<td>-0.32708</td>
<td>0.19692</td>
<td>0.54514</td>
<td>0.37819</td>
<td>-0.14720</td>
<td>0.11824</td>
<td>0.17278</td>
<td>-0.09485</td>
<td>43.38542</td>
</tr>
<tr>
<td>3000</td>
<td>-0.33788</td>
<td>0.20963</td>
<td>0.56047</td>
<td>0.38500</td>
<td>-0.14934</td>
<td>0.11979</td>
<td>0.17418</td>
<td>-0.09434</td>
<td>41.81658</td>
</tr>
</tbody>
</table>

True values -0.59000 0.45000 0.90000 0.62000 -0.25000 0.27000 0.30000 -0.20000

Table 2. The FF-SG estimates and errors.

<table>
<thead>
<tr>
<th>t</th>
<th>$a_1$</th>
<th>$a_2$</th>
<th>$b_1$</th>
<th>$b_2$</th>
<th>$c_1$</th>
<th>$c_2$</th>
<th>$d_1$</th>
<th>$d_2$</th>
<th>$\delta$(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>-0.26954</td>
<td>0.12798</td>
<td>0.44137</td>
<td>0.34673</td>
<td>-0.14464</td>
<td>0.07554</td>
<td>0.16169</td>
<td>-0.05871</td>
<td>53.68516</td>
</tr>
<tr>
<td>200</td>
<td>-0.28349</td>
<td>0.16094</td>
<td>0.50401</td>
<td>0.37156</td>
<td>-0.14528</td>
<td>0.09701</td>
<td>0.16561</td>
<td>-0.07652</td>
<td>48.36300</td>
</tr>
<tr>
<td>500</td>
<td>-0.35597</td>
<td>0.21982</td>
<td>0.61865</td>
<td>0.41566</td>
<td>-0.17942</td>
<td>0.14936</td>
<td>0.20513</td>
<td>-0.11309</td>
<td>36.28997</td>
</tr>
<tr>
<td>1000</td>
<td>-0.46405</td>
<td>0.33428</td>
<td>0.74986</td>
<td>0.47299</td>
<td>-0.16999</td>
<td>0.19877</td>
<td>0.21277</td>
<td>-0.15816</td>
<td>21.65130</td>
</tr>
<tr>
<td>2000</td>
<td>-0.57771</td>
<td>0.41033</td>
<td>0.87268</td>
<td>0.52780</td>
<td>-0.21269</td>
<td>0.23785</td>
<td>0.26543</td>
<td>-0.19347</td>
<td>8.53179</td>
</tr>
<tr>
<td>3000</td>
<td>-0.60217</td>
<td>0.43336</td>
<td>0.90331</td>
<td>0.59785</td>
<td>-0.25448</td>
<td>0.25998</td>
<td>0.29719</td>
<td>-0.18394</td>
<td>2.55572</td>
</tr>
</tbody>
</table>

True values -0.59000 0.45000 0.90000 0.62000 -0.25000 0.27000 0.30000 -0.20000

Table 3. The TVFF-SG estimates and errors.

<table>
<thead>
<tr>
<th>t</th>
<th>$a_1$</th>
<th>$a_2$</th>
<th>$b_1$</th>
<th>$b_2$</th>
<th>$c_1$</th>
<th>$c_2$</th>
<th>$d_1$</th>
<th>$d_2$</th>
<th>$\delta$(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>-0.27397</td>
<td>0.13478</td>
<td>0.45040</td>
<td>0.35132</td>
<td>-0.14404</td>
<td>0.07715</td>
<td>0.16166</td>
<td>-0.05979</td>
<td>52.81492</td>
</tr>
<tr>
<td>200</td>
<td>-0.29024</td>
<td>0.17352</td>
<td>0.52379</td>
<td>0.38015</td>
<td>-0.14544</td>
<td>0.10261</td>
<td>0.16697</td>
<td>-0.08105</td>
<td>46.57833</td>
</tr>
<tr>
<td>500</td>
<td>-0.37904</td>
<td>0.24131</td>
<td>0.65674</td>
<td>0.43346</td>
<td>-0.18747</td>
<td>0.16561</td>
<td>0.21458</td>
<td>-0.12435</td>
<td>32.30918</td>
</tr>
<tr>
<td>1000</td>
<td>-0.50125</td>
<td>0.36779</td>
<td>0.79887</td>
<td>0.50379</td>
<td>-0.18071</td>
<td>0.22027</td>
<td>0.22821</td>
<td>-0.17431</td>
<td>15.99839</td>
</tr>
<tr>
<td>2000</td>
<td>-0.60015</td>
<td>0.42432</td>
<td>0.89935</td>
<td>0.55162</td>
<td>-0.22631</td>
<td>0.25071</td>
<td>0.28062</td>
<td>-0.20563</td>
<td>5.80706</td>
</tr>
<tr>
<td>3000</td>
<td>-0.60405</td>
<td>0.44296</td>
<td>0.90256</td>
<td>0.61737</td>
<td>-0.24607</td>
<td>0.26640</td>
<td>0.29327</td>
<td>-0.19900</td>
<td>1.29122</td>
</tr>
</tbody>
</table>

True values -0.59000 0.45000 0.90000 0.62000 -0.25000 0.27000 0.30000 -0.20000
Figure 1. The parameter estimation errors $\delta$ vs. $t$.

Figure 2. The time evolution curves of the system parameters vs. $t$. 
tially supported by the National Natural Science Foundation of China (Grant No.11771376) and the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (Grant No.13KJD520010).

REFERENCES