International Journal of Innovative Computing, Information and Control
Volume **12**, Number **3**, June **2016**

ICIC International ©2016 ISSN 1349-4198
pp. 747–759

THE IMPROVED STOCHASTIC NEWTON ALGORITHM FOR DUAL-RATE SYSTEM IDENTIFICATION

Pengfei Cao[1,2] and Xionglin Luo[2]

[1]College of Electrical Engineering and Automation
Shandong University of Science and Technology
No. 579, Qianwangang Road, Qingdao 266590, P. R. China
cpfskdzdh@sina.com

[2]Research Institute of Automation
China University of Petroleum
No. 18, Fuxue Road, Changping District, Beijing 102249, P. R. China
luoxl@cup.edu.cn

Received November 2015; revised March 2016

Abstract. *The stochastic Newton recursive algorithm is studied and utilized for identifying the dual-rate model which is derived through the polynomial transformation technique. The main advantage of the algorithm is that it has the extensive form and may earn more possible convergence properties with flexible parameters. The primary problem is that the sample covariance matrix may be singular when encountering numbers of model parameters and (or) no general input signal, which hinders the identification process. Thus, the main contribution is adding one symmetric positive definite matrix to the recursion of the sample covariance matrix. This simple approach solves the problem effectively. Two improved stochastic Newton recursive algorithms are then proposed for time-invariant and time-varying systems. The consistent and bounded convergence conclusions of the corresponding algorithms are drawn respectively. The final illustrative examples demonstrate the effectiveness and the convergence properties of the recursive algorithms.*
Keywords: Dual-rate system, System identification, Recursive algorithm, Consistent convergence, Bounded convergence

1. **Introduction.** A system with two sampling frequencies is called dual-rate system, in which the ones with slow-sampled outputs and fast-sampled inputs widely exist in chemical processes [1-3]. In this study, the single input and single output (SISO) dual-rate system is investigated.

Studies on dual-rate systems have been undertaken actively in terms of both theory and applications, e.g., system modelling and identification [3,4], predictive control [5], and fault detection [6]. Modelling and identification are the primary tasks for dual-rate systems [7,8]. However, scarce measurements of outputs bring in difficulties for model identification. Accurate model cannot be obtained just using conventional approaches. A number of main methods have been developed to obtain smooth intersample predictions and then an appropriate single-rate model [9,10]. The polynomial transformation technique is one of the main and well-developed methods. The dual-rate model is derived through this technique and could reveal just the relation between input and output at sampling time. Unmeasurable outputs would not appear in the model. After identifying the dual-rate model, the single-rate one could be calculated following the inverse process with the technique. Therefore, we choose this technique to obtain a dual-rate model first and identify it with our algorithm [3,11-14].

The stochastic Newton recursive (SNR) algorithm is adopted, and the improvement research on it is mainly studied for dual-rate model identification. The SNR algorithm is based on the concept of gradient descent and employs a sample covariance matrix to control update direction [15,16]. With extensive form and flexible parameters, the SNR algorithm may be performed to adapt to different changes in production requirements and working conditions. Its description involves two update relations: for updating the weight estimate and the sample covariance matrix. However, when too many model parameters are encountered or the input signal is not sufficiently general, the covariance matrix may be nearly singular and the "ill-posed" problem of performing covariance matrix inversion occurs [17,18], which hinder the identification process. This problem may also appear in many other algorithms. The discussed reasons above are easy to meet. For example, the numbers of the parameters of the derived dual-rate model are multiplied compared with the single-one, especially with multiple variables in the system; actual process is stable in general, and it is hard to obtain persistently exciting input signal. In numbers of simulation experiments of our last researches [19-21], the ill-posed problem arises repeatedly and the parameter estimations are divergent. Ljung and Soderstron suggested adding one diagonal positive definite matrix to the covariance matrix [19]. It indeed worked in our tests. Actually, more matrixes could be considered to guarantee the covariance matrix positive definite. However, there is lack of research on the convergence analysis of the improved algorithm and the conditions the added matrix should satisfy. Therefore, in this work, we mainly try to add one specific symmetric positive definite matrix to the covariance matrix, and study the convergence of the improved SNR algorithm and give the conditions the added matrix should meet. Two improved SNR algorithms are proposed for time-invariant and time-varying systems. Then, the convergence conclusions are drawn, respectively. For dual-rate time-invariant systems, the parameter estimations would be consistently convergent to the true values based on the corresponding algorithm. When the system is time-varying and the change rules for the parameters are unknown, the consistent convergence of the model parameters does not exist [22]. However, the parameter estimation errors will be mean square bounded based on the corresponding algorithm. The bounded convergence conclusion indicates good tracking performance of the algorithm.

The rest of the paper is organized as follows. In Section 2, the dual-rate model and the improved SNR algorithms are proposed. In Section 3, the consistent and bounded convergences are demonstrated respectively for these two algorithms. Section 4 provides two examples to illustrate the effectiveness of the algorithms and the convergence conclusions. The final remarks are given in Section 5.

2. Dual-Rate Model and the Improved SNR Algorithms.

In this section, the dual-rate model is firstly introduced. Then, two improved SNR algorithms are proposed for dual-rate time-invariant and time-varying systems.

2.1. Dual-rate model.

The following single-rate model below can be utilized to represent a dual-rate system [11]:

$$A(z)y(k) = B(z)u(k)$$

$$A(z) = 1 + a_1 z^{-1} + \cdots + a_n z^{-n} = \prod_{i=1}^{n} (1 - z_i z^{-1}), \quad B(z) = b_0 + b_1 z^{-1} + \cdots + b_n z^{-n}$$

where $\{a_i\}$ and $\{b_i\}$ are the model parameters. The roots of $A(z)$ are assumed as $\{z_i\}$. We define z as the forward shift operator, and $z^{-1}x(k) = x(k-1)$. For this dual-rate system, the sampling period of the output y is assumed as q times that of the input u (q

is a positive integer). Given that the intersample data $\{y(kq+i),\ i=1,2,\ldots,(q-1)\}$ cannot be sampled, this single-rate model would not be identified directly.

Therefore, the polynomial transformation technique is used to obtain the dual-rate model with noise term $\upsilon(kq)$ as follows [3,11]:

$$\alpha(z)y(kq) = \beta(z)u(kq) + \upsilon(kq)$$
$$\alpha(z) = \gamma(z)A(z) = 1 + \alpha_1 z^{-q} + \alpha_2 z^{-2q} + \cdots + \alpha_n z^{-nq}$$
$$\beta(z) = \gamma(z)B(z) = \beta_1 z^{-1} + \cdots + \beta_n z^{-nq}$$
$$\gamma(z) = \prod_{i=1}^{n}\left(1 + z_i z^{-1} + z_i^2 z^{-2} + \cdots + z_i^{q-1} z^{-q+1}\right)$$

where $\{\alpha_i\}$ and $\{\beta_i\}$ are the dual-rate model parameters. We obtain the following regression form:

$$y(kq) = \varphi^{\tau}(kq)\theta(kq) + \upsilon(kq) \tag{1}$$
$$\varphi(kq) = [-y(kq-q), -y(kq-2q), \ldots, -y(kq-nq), u(kq-1), \ldots, u(kq-nq)]^{\tau}$$
$$\theta(kq) = [\alpha_1, \alpha_2, \ldots, \alpha_n, \beta_1, \beta_2, \ldots, \beta_{nq}]^{\tau}$$

After deriving the dual-rate model, i.e., Equation (1), the relation between input and output at fixed relative time positions is determined. This derived dual-rate model can be identified only using available data, which exhibits the main advantage of the polynomial transformation technique. In addition, the linear periodically time-varying characteristic would never appear [7,23]. For time-invariant single-rate systems, dual-rate model parameters are also time-invariant. Actually, this regression structure in Equation (1) could also be used for representing certain nonlinear systems. Therefore, the following algorithms could also be utilized for nonlinear system identification.

2.2. **Improved SNR algorithms.** The SNR algorithm is shown as follows [15,16]:

$$\begin{cases} \hat{\theta}(kq) = \hat{\theta}(kq-q) + \rho(kq)R^{-1}(kq)\varphi(kq)\left[y(kq) - \varphi^{\tau}(kq)\hat{\theta}(kq)\right] \\ R(kq) = R(kq-q) + \rho(kq)\left[\varphi(kq)\varphi^{\tau}(kq) - R(kq-q)\right] \end{cases}$$

where $R(kq)$ represents the sample covariance matrix and $\rho(kq)$ represents the forgetting factor. Actually, the first equation of the algorithm is the main part and $R(kq)$ could be substituted with some other constant matrixes. In addition, no fixed rules are arranged for $\{\rho(kq)\}$, and the forgetting factor would make adjustments in real-time with working condition changing. This algorithm could be converted into many other ones with certain values for $R(kq)$ and $\{\rho(kq)\}$. Therefore, the SNR algorithm has extensive form and embraces more possible performance. However, $R(kq)$ may be singular when too many parameters occur and (or) the input signal is not general. The "ill-posed" problem may appear when performing the inversion of nearly singular $R(kq)$, and this hinders the identification for $\hat{\theta}(kq)$ [18]. This situation may also appear for many other recursive algorithms. To address this problem, we add a symmetric positive definite matrix to the covariance matrix to ensure that it will remain positive definite. The improved SNR algorithm for dual-rate time-invariant model is presented as follows:

$$\begin{cases} \hat{\theta}(kq) = \hat{\theta}(kq-q) + \rho(kq)R^{-1}(kq)\varphi(kq)\left[y(kq) - \varphi^{\tau}(kq)\hat{\theta}(kq)\right] \\ R(kq) = R(kq-q) + \rho(kq)\left[\varphi(kq)\varphi^{\tau}(kq) + A(kq) - R(kq-q)\right] \end{cases} \tag{2}$$

where $A(kq)$ is the symmetric positive definite matrix. It has been proved that $R(kq)$ would always keep positive definite (in Section 3).

For dual-rate time-invariant systems, the determination of $A(kq)$ should follow Equation (A.6) at least to ensure the convergency of the parameter estimations. However, for time-varying systems, we do not exactly know how the parameters change. It is unable to find one algorithm to eliminate the estimation errors thoroughly. So, we have proposed the algorithm (Equation (3)) to track the parameter changes as much as possible and guarantee the bounded convergence of the algorithm. Considering the convenience for applications at the same time, it is enough to set one constant matrix for $A(kq)$ (following A5 in Section 3.2). And the corresponding recursive algorithm for time-varying system is shown as:

$$\begin{cases} \hat{\theta}(kq) = \hat{\theta}(kq-q) + \lambda R^{-1}(kq)\varphi(kq)\left[y(kq) - \varphi^{\tau}(kq)\hat{\theta}(kq)\right] \\ R(kq) = R(kq-q) + \lambda\left[\varphi(kq)\varphi^{\tau}(kq) + A - R(kq-q)\right] \end{cases} \tag{3}$$

where constant λ $(0 < \lambda < 1)$ has been chosen as forgetting factor. The reason for such arrangement lies in that $\rho(kq)R^{-1}(kq)\varphi(kq)$ approaches to zero as k increases according to Equation (2) and Theorem 3.1. However, the constant forgetting factor could guarantee the tracking performance of the algorithm. Determinations for $A(kq)$, A, $\{\rho(kq)\}$ and λ will be discussed in Section 3.

3. **Convergence Analysis of the Improved SNR Algorithms.** In this section, two convergence conclusions are drawn for the improved algorithms respectively.

3.1. **Consistent convergence of the algorithm.** We define the true dual-rate model parameter vector as θ for time-invariant systems. Hence, we have the consistent convergence (shown in Theorem 3.1) for $\hat{\theta}(kq)$ based on the Martingale Convergence Theorem.

Lemma 3.1. (Martingale Convergence Theorem [24,25]). *Assume that $\{W(t)\}$, $\{f(t)\}$ and $\{g(t)\}$, which are the non-negative random variable sequences and adapted to the σ-algebra sequence $\{F_t\}$, satisfy the following relation:*

$$E[W(t)|F_t] \le W(t-1) - f(t) + g(t)$$

If $\sum_{t=1}^{\infty} g(t) < \infty$, a.s., then $W(t)$ a.s. converges to a finite random variable W_0, i.e., $W(t) \to W_0$, a.s. and $\sum_{t=1}^{\infty} f(t) < \infty$, a.s.

Theorem 3.1. *Assume that conditions A1-A3 hold.*

A1 $\{F_{kq-q}\}$ is the σ-algebra sequence [11,26], and the noise sequence $\{\upsilon(kq)\}$ satisfies

$$E\left[\upsilon(kq)|F_{kq-q}\right] = 0, \quad E\left[\upsilon^2(kq)|F_{kq-q}\right] = \sigma_\upsilon^2(kq) \le \sigma_\upsilon^2 < \infty$$

A2 The generalized persistent excitation condition holds, and information vector $\varphi(kq)$ satisfies

$$\alpha I \le \frac{1}{N}\sum_{i=1}^{N} \varphi(kq-iq)\varphi^{\tau}(kq-iq) \le \gamma I\ ,\ N > 0,\ 0 < \alpha \le \gamma < \infty$$

$$0 \le \|\varphi(kq)\|^2 \le M < \infty$$

A3 The forgetting factor $\{\rho(kq)\}$ satisfies [18,24]

$$0 < \rho(kq) < 1, \quad \sum_{k=1}^{\infty} \rho(kq) = \infty, \quad \sum_{k=1}^{\infty} \rho^2(kq) < \infty$$

Then, $\hat{\theta}(kq)$ will be consistently convergent to θ in final, that is $\lim_{k\to\infty} \hat{\theta}(kq) = \theta$. The proof for Theorem 3.1 refers to Appendix A.

Remark 3.1. *The recursive algorithm exhibits the property of fast convergence and a strong capability to resist noise. However, when the system is time-varying, the algorithm is inapplicable and the parameter changes cannot be tracked.*

Remark 3.2. *During the proof of Theorem 3.1, we only discuss the existence of $A(kq)$ which is difficult to determine. The diagonal matrix $\lambda_1(kq)I$ (I is unit diagonal matrix) is usually substituted for $A(kq)$ in applications. According to Equation (A.6), $\lambda_1(kq)$ can be determined based on the following equation:*

$$\frac{\lambda_1(kq)}{\lambda_{\min}(R(kq-q))} = 1 \tag{4}$$

3.2. **Bounded convergence of the algorithm.** For dual-rate time-varying systems, the consistent convergence for the parameter estimations does not exist [22]. In this study, we prove that $\left\|\hat{\theta}(kq) - \theta(kq)\right\|^2$ will be mean square bounded (Theorem 3.2) based on Martingale Hyperconvergence Theorem.

Lemma 3.2. (Martingale Hyperconvergence Theorem [27,28]). *Assume the following non-negative definite function*

$$T(t) = T[x(t)] = \|x(t)\|^2$$

and collection set

$$R_t = [x(t) : g[x(t)] \le \eta_t < \infty,\ a.s.]$$

where $g(x) = (a^x x)^2$ denotes the terms for the convergence variable, and a is a non-zero time-variant or time-invariant vector. η_t ($\eta_t \ge 0$) is a non-reduced and bounded stochastic variable. $b(t)$ is a stochastic variable and $(x(t), F_t)$ is an adaptive sequence. If the following equation below holds when $x(t) \in R_t^c$:

$$E[T(t+1)|F_t] - T(t) \le -b(t+1). \quad a.s.$$

where R_t^c is the complementary set of R_t, then $\lim_{t\to\infty} x(t) \in R_t$ if $b(t) \ge \frac{b}{t}$, $b > 0$.

Theorem 3.2. *Assume the conditions A1, A2, A4 and A5 hold.*

A1 and A2 are the same as the ones in Theorem 3.1.

A4 The parameter change $w(kq) = \theta(kq) - \theta(kq-q)$ is uncorrelated and mean square bounded, namely,

$$E[w(tq)w^\tau(sq)] = 0,\ t \ne s, \quad E\left[\|w(kq)\|^2\right] = \sigma_w^2(kq) \le \sigma_w^2 < \infty$$

$$E[v(tq)w(sq)] = 0$$

The change period is h-times (h is some positive integer and satisfies $h > 1$) the sampling period of $y(kq)$; hence, $w(kq)$ satisfies

$$w(kq) = \begin{cases} w(kq) & k = nhq \\ 0, & other \end{cases}$$

A5 A is a symmetric positive definite matrix. R_0 is the initial value of $R(kq)$. A and R_0 satisfy

$$R_0 = P^{-1}\Lambda P, \quad R_0 > A, \quad \lambda_{\max}(A) - \lambda_{\min}(A) \le \lambda\alpha$$

where P is the eigenvector matrix of A and Λ is the diagonal matrix that consists of the eigenvalues of R_0.

Then, $\tilde{\theta}(kq) = \hat{\theta}(kq) - \theta(kq)$ will be mean square bounded, that is,

$$\lim_{k\to\infty} E\left\|\tilde{\theta}(kq)\right\|^2 \le \frac{\lambda^2\sigma_v^2 M + \alpha(\gamma + \lambda_{\max}(A))\sigma_w^2}{\alpha\lambda\left[\alpha - (\lambda_{\max}(A) - \lambda_{\min}(A))\right]}$$

The proof for Theorem 3.2 refers to Appendix B.

Remark 3.3. *The bounded convergence indicates that the algorithm of Equation (3) exhibits good tracking performance for dual-rate time-varying systems. However,* $\lambda R^{-1}(kq)$ $\varphi(kq)\left[y(kq)-\varphi^{\tau}(kq)\hat{\theta}(kq-q)\right]$ *will never converge to zero due to the existence of noise, and this algorithm may become invalid for time-invariant systems.*

Remark 3.4. *In general, A is difficult to determine. Therefore, we substitute diagonal matrix $\lambda_1 I$ for A in applications. λ_1 is easy to determine according to A5. The estimation error bound of the dual-rate model parameters can be described as follows:*

$$\lim_{k\to\infty} E\left\|\tilde{\theta}(kq)\right\|^2 \le \frac{\lambda^2\sigma_v^2 M + \alpha\left(\gamma+\lambda_1\right)\sigma_w^2}{\alpha^2\lambda}$$

Remark 3.5. *The given parameter change rule is the typical one that frequently agrees with actual industrial processes. The bounded convergence conclusion of the recursive algorithm is correct in this case. If the parameters change in another way, the parameter estimation error will increase or the estimations will diverge. In our further study, we will improve the recursive algorithm to ensure the bounded convergence for the parameter estimations with more changing rules.*

4. Case Study.

4.1. Example 1.

Consider a single-rate time-invariant system with

$$A(z) = 1 - a_1 z^{-1} + a_2 z^{-2} = 1 - 1.6z^{-1} + 0.8z^{-2}$$

$$B(z) = b_1 z^{-1} + b_2 z^{-2} = 0.412z^{-1} + 0.309z^{-2}$$

The simulation experiment is conducted for 100s. Take the sampling period of input as $T = 1$s and $q = 2$. Let $\gamma(z) = 1 + 1.6z^{-1} + 0.8z^{-2}$. We derive the dual-rate model with a noise term as follows:

$$y(kq) = \varphi^{\tau}(kq)\theta(kq) + v(kq)$$

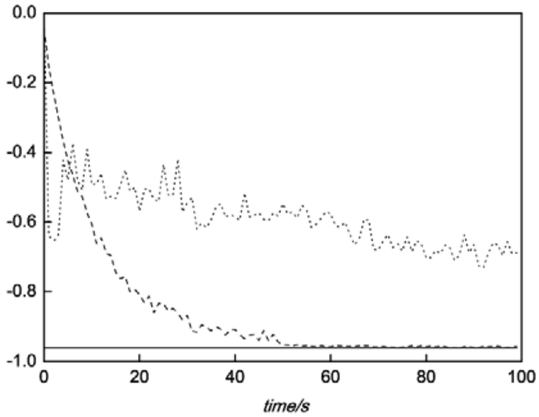
$$\varphi(kq) = [y(kq-2), y(kq-4), u(kq-1), u(kq-2), u(kq-3), u(kq-4)]^{\tau}$$

$$\theta = [-0.96, -0.64, 0.412, 0.9682, 0.824, 0.2472]^{\tau}$$

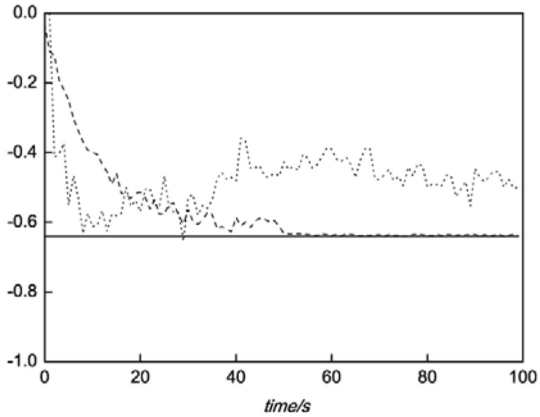
We define the parameters of the dual-rate model $\theta = [\alpha_1, \alpha_2, \beta_1, \beta_2, \beta_3, \beta_4]^{\tau}$, which will be identified with the improved SNR algorithm in Equation (2).

The initial values are set as: $\hat{\theta}(0) = 0_{6\times 1}$, and $R(0) = \eta^2 I_{6\times 6}$, where η is some large constant. The forgetting factor $\rho(kq) = \frac{1}{k}$ and $\lambda_1(kq)$ is determined following Equation (4). As one effective and widely used method, the stochastic gradient (SG) algorithm is also adopted for the model identification [11,13,14]. The parameter estimations $\left\{\hat{\theta}(kq), \hat{\theta}(kq+i)|\hat{\theta}(kq+i) = \hat{\theta}(kq), i = 1, 2, \ldots, q-1, k = 1, 2, \ldots\right\}$ from these two algorithms are shown in Figure 1.

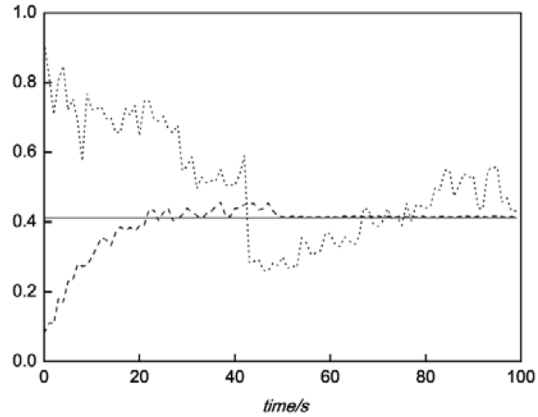
As shown in Figure 1, the parameter estimations are consistently convergent to the true values finally based on the improved SNR algorithm. On the contrary, the estimations from the SG algorithm show slow convergence speed. It has been illustrated that based on SG algorithm the parameter estimations would converge to the true values in final, and it will take quite a long time [11,13]. The improved SNR algorithm shows faster convergence. We have also tried the SNR algorithm without the added matrix. The result is that the sample covariance matrix is indeed singular and the parameter estimations diverge, which confirm the necessity of the improvement in this paper.



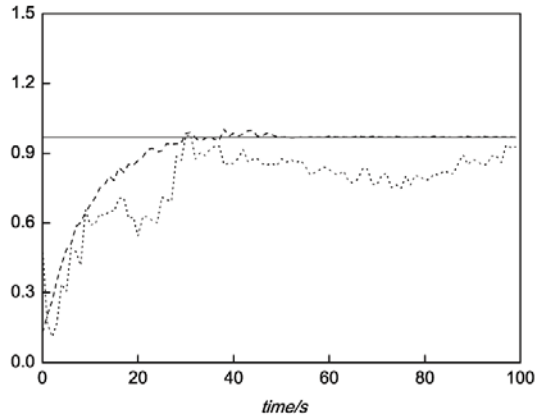
(a) Raw values and the estimations of α_1



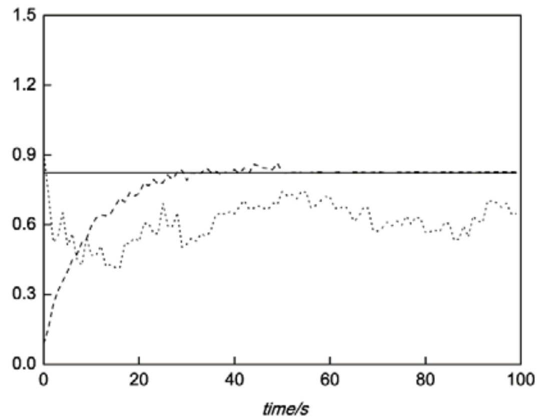
(b) Raw values and the estimations of α_2



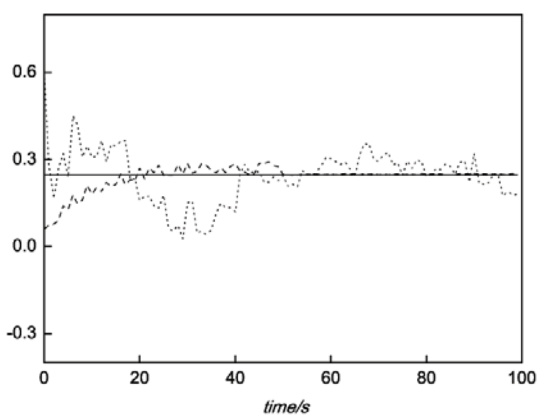
(c) Raw values and the estimations of β_1



(d) Raw values and the estimations of β_2



(e) Raw values and the estimations of β_3

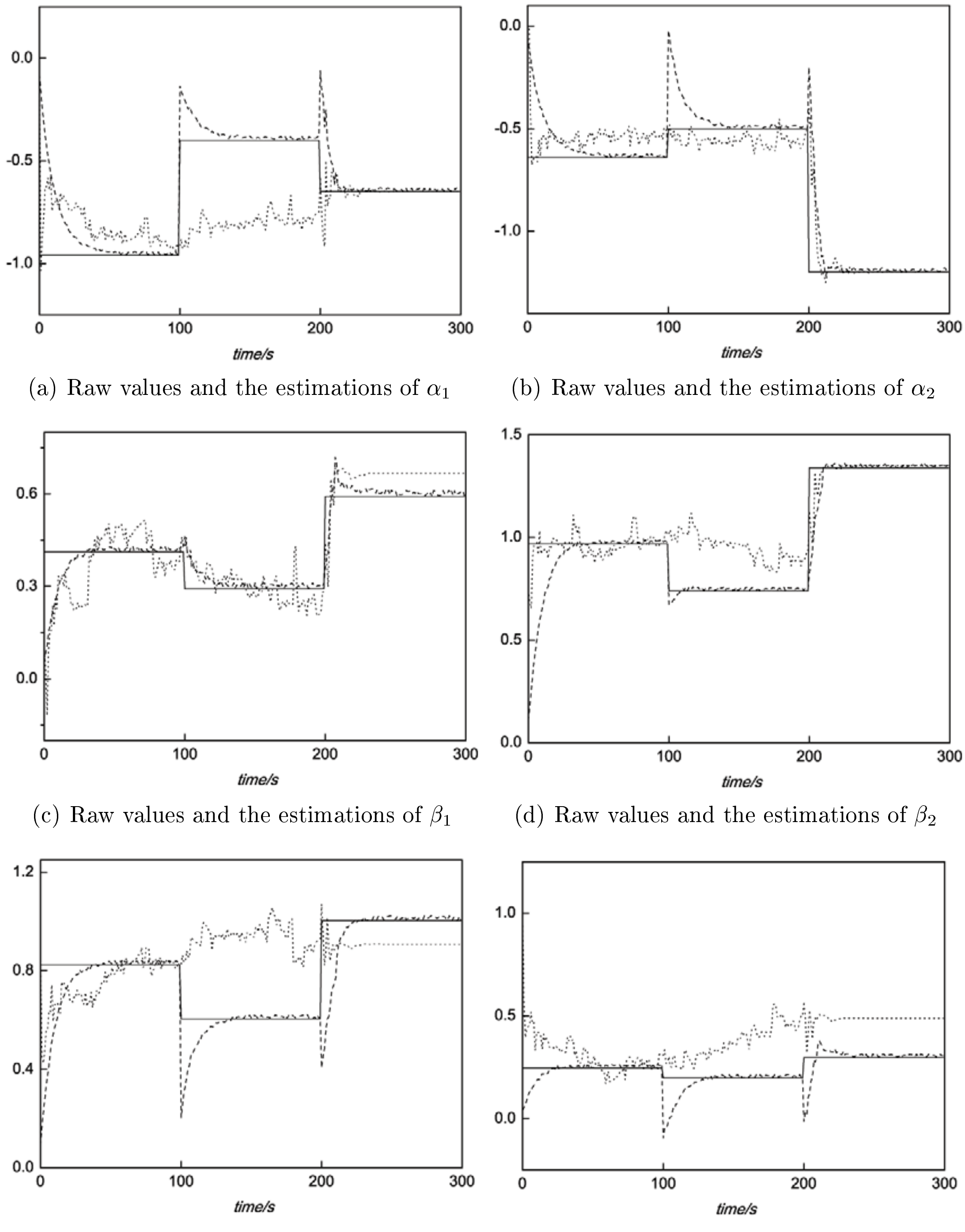


(f) Raw values and the estimations of β_4

FIGURE 1. Raw values (solid line) and estimations of the dual-rate model parameters. $---$ from the proposed algorithm; $\cdots$ from SG algorithm

4.2. **Example 2.** The dual-rate model in Example 1 with time-varying parameters is identified based on the algorithm of Equation (3).

The experiment is conducted for 300s, and T and q are set as the same values in Example 1. We set the parameter change period as 50 times that of the sampling period of the output. The initial values are set as: $\hat{\theta}(0) = 0_{6\times 1}$, and $R(0) = \eta^2 I_{6\times 6}$, where η is a large constant. We also set $\lambda = 0.5$ and $\lambda_1 = 0.05$. The SG algorithm with forgetting factor is also adopted for model identification. The parameter estimations



(a) Raw values and the estimations of α_1 (b) Raw values and the estimations of α_2

(c) Raw values and the estimations of β_1 (d) Raw values and the estimations of β_2

(e) Raw values and the estimations of β_3 (f) Raw values and the estimations of β_4

FIGURE 2. Raw values (solid line) and estimations of the dual-rate model parameters. $---$from the proposed algorithm; $\cdots$ from SG algorithm

$\left\{\hat{\theta}(kq), \hat{\theta}(kq+i) | \hat{\theta}(kq+i) = \hat{\theta}(kq), i = 1, 2, \ldots, q-1, k = 1, 2, \ldots\right\}$ are shown in Figure 2.

As evident from Figure 2, when the model parameters change, the estimations would be adjusted rapidly to track the changes using the improved algorithm. And between every two changing moments, the parameter estimation errors are controlled within smaller range. Therefore, the parameter estimations always remain boundedly convergent. However, the estimation errors from the SG algorithm may be relatively large. The effectiveness of the algorithm has been illustrated.

5. **Conclusions.** In this study, the SNR algorithm is studied to identify the dual-rate model. This algorithm has extensive form and could be converted into many other algorithms. In addition, the SNR algorithm may have lots of convergence properties with flexible parameters. Thus, the algorithm shows more research value. To avoid the singular sample covariance matrix, a symmetric positive definite matrix is added to the recursion of the covariance matrix, and the recursion identification can proceed. We have just provided a simple approach to solve the problem. However, there is less research on the algorithm convergence with the added matrix, as well as for many other algorithms. We have also proposed two improved SNR algorithms for dual-rate time-invariant and time-varying systems, and illustrated the corresponding convergence conclusions in detail.

We have made preliminary research on the recursive algorithm, but many issues are worth studying in the further work. For example, from Theorems 3.1 and 3.2 we know that the forgetting factor has a great deal with the convergence speed and estimation error, and there must be the optimal choices for $\{\rho(kq)\}$ and λ. Therefore, some ways should be found to determine the forgetting factor in real-time to adapt to working condition change.

Acknowledgment. This work was supported by the National Basic Research Program of China (2012CB720500).

REFERENCES

[1] E. L. Lehmann and G. Casella, *Theory of Point Estimation*, 2nd Edition, Springer, New York, 1998.

[2] Y. Wu and X. L. Luo, Novel calibration approach of soft sensor based on multirate data fusion technology, *Journal of Process Control*, vol.20, pp.1252-1260, 2010.

[3] L. Bao, R. Bodil, M. S. Torben, K. H. K. Jorgen and B. J. Sten, Data-driven soft sensor design with multiple-rate sampled data: A comparative study, *Ind. Eng. Chem. Res.*, vol.48, pp.5379-5387, 2009.

[4] X. G. Liu and J. Lu, Least squares based iterative identification for a class of multirate systems, *Automatica*, vol.46, pp.549-554, 2010.

[5] R. Harigopal, K. T. Arun, R. B. Gopaluni and S. L. Shah, Identification of chemical processes with irregular output sampling, *Control Engineering Practice*, vol.14, pp.467-480, 2006.

[6] M. Embirucu and C. Fontes, Multirate multivariable generalized predictive control and its application to a slurry reactor for ethylene polymerization, *Chemical Engineering Science*, vol.61, no.17, pp.5754-5767, 2006.

[7] W. Li, S. L. Shah and D. Y. Xiao, Kalman filters in non-uniformly sampled multirate systems: For FDI and beyond, *Automatica*, vol.44, no.1, pp.199-208, 2008.

[8] J. D. Wang, T. W. Chen and B. Huang, Multi-rate sampled-data systems: Computing fast-rate models, *Journal of Process Control*, vol.14, pp.79-88, 2004.

[9] Y. C. Zhu, H. Yelkamp, J. H. Wang and Q. Fu, System identification using slow and irregular output samples, *Journal of Process Control*, vol.19, pp.58-67, 2009.

[10] R. Amirhalingam and J. H. Lee, Subspace identification based inferential control applied to a continuous pulp digester, *Journal of Process Control*, vol.9, no.5, pp.397-406, 1995.

[11] P. Ramachandran, G. E. Young and E. A. Misawa, Intersample output estimation with multirate sampling, *Proc. of IEEE Int. Conf. Control Appl.*, vol.15-18, pp.576-581, 1996.

[12] F. Ding, P. X. Liu and Y. Shi, Convergence analysis of estimation algorithms for dual-rate stochastic systems, *Applied Mathematics and Computation*, vol.176, pp.245-261, 2006.

[13] J. Ding and F. Ding, The residual based extended least squares identification method for dual-rate systems, *Computers and Mathematics with Applications*, vol.56, pp.1479-1487, 2008.

[14] J. Ding, S. Yang, H. G. Wang and F. Ding, A modified stochastic gradient based parameter estimation algorithm for dual-rate sampled data systems, *Digital Signal Processing*, vol.20, pp.1238-1247, 2010.

[15] F. Ding and T. W. Chen, A gradient based adaptive control algorithm for dual-rate systems, *Asian Journal of Control*, vol.8, no.4, pp.1-10, 2006.
[16] O. P. Ferreira, M. L. N. Goncalves and P. R. Oliveira, Local convergence analysis of inexact Gauss-Newton like methods under majorant condition, *Journal of Computational and Applied Mathematics*, vol.236, no.9, pp.2487-2498, 2012.
[17] A. D. Richard, Gauss-Newton and M-estimation for ARMA processes with infinite variance, *Stochastic Processes and Their Applications*, vol.63, pp.75-95, 1996.
[18] T. Tarvainen, M. Vauhkonen and S. R. Arridge, Gauss-Newton reconstruction method for optical tomography using the finite element solution of the radiative transfer equation, *Journal of Quantitative Spectroscopy and Radiative Transfer*, vol.109, nos.17-18, pp.2767-2778, 2008.
[19] L. Ljung and T. Soderstron, *Theory and Practice of Recursive Identification*, MIT Press, 1983.
[20] P. F. Cao and X. L. Luo, Auxiliary-model-based stochastic Newton recursive identification for dual-rate system, *Journal of Control and Decision*, vol.1, pp.117-124, 2015.
[21] P. F. Cao and X. L. Luo, Wiener structure based modeling and identifying of soft sensor systems, *Acta Automatica Sinica*, vol.10, pp.2179-2192, 2014.
[22] P. F. Cao and X. L. Luo, Modeling for soft sensor systems and parameters updating online, *Journal of Process Control*, vol.24, no.6, pp.975-990, 2014.
[23] V. Solo, Stochastic adaptive control and martingale limit theory, *IEEE Trans. Automat. Control*, vol.35, no.1, pp.66-71, 1999.
[24] G. M. Lay, Input-output analysis of multirate feedback systems, *IEEE Trans. Automat. Control*, vol.3, pp.21-28, 1957.
[25] J. Ding and F. Ding, The residual based extended least squares identification method for dual-rate systems, *Computers and Mathematics with Applications*, vol.56, pp.1479-1487, 2008.
[26] G. C. Goodwin and K. S. Sin, *Adaptive Filtering Prediction and Control*, Prentice-Hall Press, Englewood Cliffs, NJ, 1984.
[27] F. Ding, Martingale hyperconvergence theorem and the convergence of forgetting factor least squares algorithm, *Control Theory and Applications*, vol.14, no.1, pp.90-95, 1997.
[28] F. Ding, J. B. Yang and Y. M. Xu, Convergence analysis of forgetting gradient algorithm by using martingale hyperconvergence theorem, *Tsinghua Science and Technology*, vol.5, no.2, pp.187-192, 2000.

Appendix A. **Proof for Theorem 3.1.** We can get from Equation (2)

$$\begin{aligned}\tilde{\theta}(kq) &= \hat{\theta}(kq) - \theta \\ &= \tilde{\theta}(kq-q) + \rho(kq)R^{-1}(kq)\varphi(kq)\left[y(kq) - \varphi^{\tau}(kq)\hat{\theta}(kq-q)\right] \\ &= \tilde{\theta}(kq-q) + \rho(kq)R^{-1}(kq)\varphi(kq)\left[-\tilde{y}(kq) + \upsilon(kq)\right] \end{aligned} \tag{A.1}$$

where $\tilde{y}(kq) = \varphi^{\tau}(kq)\tilde{\theta}(kq-q)$. Let $T(kq) = \tilde{\theta}^{\tau}(kq)R(kq)\tilde{\theta}(kq)$. We can derive

$$\begin{aligned}T(kq) = &(1-\rho(kq))T(kq-q) + \rho(kq)\tilde{\theta}^{\tau}(kq-q)A(kq)\tilde{\theta}(kq-q) \\ &- \left[\rho(kq) - \rho^2(kq)\varphi^{\tau}(kq)R^{-1}(kq-q)\varphi(kq)\right]\tilde{y}^2(kq) \\ &+ \rho^2(kq)\varphi^{\tau}(kq)R^{-1}(kq)\varphi(kq)\upsilon^2(kq) \\ &+ \left[2\rho(kq) - \rho^2(kq)\varphi^{\tau}(kq)R^{-1}(kq)\varphi(kq)\right]\tilde{y}(kq)\upsilon(kq)\end{aligned} \tag{A.2}$$

Given that $\tilde{y}(kq)$, $\tilde{\theta}(kq)$, $\varphi(kq)$ and $R(kq)$ are uncorrelated to $\upsilon(kq)$, we take the conditional mathematical expectation for both sides of Equation (A.2) that corresponds to F_{kq-q} as follows:

$$\begin{aligned}E\left[T(kq)|F_{kq-q}\right] \le &\left(1 - \rho(kq) + \frac{\rho(kq)\lambda_{\max}(A(kq))}{\lambda_{\min}(R(kq-q))}\right)T(kq-q) \\ &- E\left[\left(\rho(kq) - \rho^2(kq)\varphi^{\tau}(kq)R^{-1}(kq)\varphi(kq)\right)\tilde{y}^2(kq)|F_{kq-q}\right] \\ &+ E\left[\rho^2(kq)\varphi^{\tau}(kq)R^{-1}(kq)\varphi(kq)|F_{kq-q}\right]\sigma_{\upsilon}^2\end{aligned} \tag{A.3}$$

According to Equation (2), we have

$$\begin{aligned}\varphi^{\tau}(kq)R^{-1}(kq)R(kq)\varphi(kq) &= (1-\rho(kq))\varphi^{\tau}(kq)R^{-1}(kq)R(kq-q)\varphi(kq)\\ &\quad+\rho(kq)\varphi^{\tau}(kq)R^{-1}(kq)A(kq)\varphi(kq)\\ &\quad+\rho(kq)\varphi^{\tau}(kq)R^{-1}(kq)\varphi(kq)\varphi^{\tau}(kq)\varphi(kq)\end{aligned}$$

Then,

$$\varphi^{\tau}(kq)\varphi(kq) > \rho(kq)\varphi^{\tau}(kq)R^{-1}(kq)\varphi(kq)\varphi^{\tau}(kq)\varphi(kq)$$

Hence,

$$\rho(kq)\varphi^{\tau}(kq)R^{-1}(kq)\varphi(kq) < 1$$

Thus,

$$E\left[\left(\rho(kq)-\rho^2(kq)\varphi^{\tau}(kq)R^{-1}(kq)\varphi(kq)\right)\tilde{y}^2(kq)|F_{kq-q}\right] > 0 \tag{A.4}$$

As clearly shown in Equation (2), when k is sufficiently large, $R(kq)\to R_0$, where R_0 is a symmetric positive definite matrix. According to A2 and A3, the last term of Equation (A.3) satisfies

$$\sum_{k=1}^{\infty} E\left[\rho^2(kq)\varphi^{\tau}(kq)R^{-1}(kq)\varphi(kq)|F_{kq-q}\right]\sigma_v^2 < \infty \tag{A.5}$$

$A(kq)$ should be determined satisfying the following equation at least:

$$\frac{\lambda_{\max}(A(kq))}{\lambda_{\min}(R(kq-q))} = 1 \tag{A.6}$$

when $k > k_1$ (k_1 is a positive integer). Therefore, the following equation below holds:

$$\sum_{k=1}^{\infty}\left|-\rho(kq)+\frac{\rho(kq)\lambda_{\max}(A(kq))}{\lambda_{\min}(R(kq-q))}\right| < \infty \tag{A.7}$$

Then, according to Lemma 3.1 and Equations (A.4)-(A.7), $T(kq)$ a.s. converges to a finite random variable T_0, such that

$$\lim_{k\to\infty} T(kq) = T_0 \tag{A.8a}$$

$$\sum_{k=1}^{\infty} E\left[\left(\rho(kq)-\rho^2(kq)\varphi^{\tau}(kq)R^{-1}(kq)\varphi(kq)\right)\tilde{y}^2(kq)|F_{kq-q}\right] < \infty \tag{A.8b}$$

Given that $R(kq)\to R_0$, $\lim\limits_{k\to\infty}\left\|\tilde{\theta}(kq)\right\|^2 = c < \infty$ and $\tilde{y}^2(kq) < \infty$ according to Equation (A.8), where c is a positive constant. Based on A3, we know that the following equation below holds:

$$\sum_{k=1}^{\infty} E\left[\left(\rho^2(kq)\varphi^{\tau}(kq)R^{-1}(kq)\varphi(kq)\right)\tilde{y}^2(kq)|F_{kq-q}\right] < \infty$$

$\sum\limits_{k=1}^{\infty} E\left[\rho(kq)\tilde{y}^2(kq)|F_{kq-q}\right] < \infty$. For that $\sum\limits_{k=1}^{\infty}\rho(kq) = \infty$, when k is sufficiently large there must be $\tilde{y}^2(kq) = 0$. Correspondingly, $\left\|\tilde{\theta}(kq)\right\|^2 = c = 0$, such that $\lim\limits_{k\to\infty}\hat{\theta}(kq) = \theta$.

The proof of Theorem 3.1 is completed.

Appendix B. **Proof for Theorem 3.2.** We can get from Equation (3)

$$\begin{aligned}\tilde{\theta}(kq) &= \hat{\theta}(kq) - \theta(kq)\\ &= \hat{\theta}(kq) - [\theta(kq-q) + w(kq)]\\ &= \tilde{\theta}(kq-q) + \lambda R^{-1}(kq)\varphi(kq)\left[-\tilde{y}(kq) + \upsilon(kq)\right] - w(kq)\end{aligned}$$

where $\tilde{y}(kq) = \varphi^{\tau}(kq)\tilde{\theta}(kq-q)$. Let $T(kq) = \tilde{\theta}^{\tau}(kq)R(kq)\tilde{\theta}(kq)$. Then, we can derive

$$\begin{aligned}T(kq) = {}& (1-\lambda)T(kq-q) + \lambda\tilde{\theta}^{\tau}(kq-q)A\tilde{\theta}(kq-q)\\ &- \left[\lambda - \lambda^2\varphi^{\tau}(kq)R^{-1}(kq)\varphi(kq)\right]\tilde{y}^2(kq)\\ &+ \lambda^2\varphi^{\tau}(kq)R^{-1}(kq)\varphi(kq)\upsilon^2(kq) + w^{\tau}(kq)R(kq)w(kq)\\ &- \left[2\lambda - 2\lambda^2\varphi^{\tau}(kq)R^{-1}(kq)\varphi(kq)\right]\tilde{y}(kq)\upsilon(kq)\\ &+ \left[-\lambda\varphi^{\tau}(kq)\tilde{y}(kq) - \tilde{\theta}^{\tau}(kq-q)R(kq) - \lambda\varphi^{\tau}(kq)\upsilon(kq)\right]w(kq)\\ &+ w^{\tau}(kq)\left[-\lambda\varphi(kq)\tilde{y}(kq) - R(kq)\tilde{\theta}(kq-q) - \lambda\varphi(kq)\upsilon(kq)\right] \qquad \text{(B.1)}\end{aligned}$$

Given that $\tilde{y}(kq)$, $\tilde{\theta}(kq)$, $\varphi(kq)$ and $R(kq)$ are uncorrelated to $\upsilon(kq)$ and $w(kq)$, we take the conditional mathematical expectation that corresponds to F_{kq-q} for both sides of Equation (B.1) as follows:

$$\begin{aligned}E\left[T(kq)|F_{kq-q}\right] \le {}& (1-\lambda)T(kq-q) + E\left[\lambda\tilde{\theta}^{\tau}(kq-q)A\tilde{\theta}(kq-q)|F_{kq-q}\right]\\ &- E\left[\left(\lambda - \lambda^2\varphi^{\tau}(kq)R^{-1}(kq)\varphi(kq)\right)\tilde{y}^2(kq)|F_{kq-q}\right]\\ &+ \lambda^2 E\left[\varphi^{\tau}(kq)R^{-1}(kq)\varphi(kq)|F_{kq-q}\right]\sigma_{\upsilon}^2\\ &+ E\left[w^{\tau}(kq)R(kq)w(kq)|F_{kq-q}\right] \qquad \text{(B.2)}\end{aligned}$$

Given that $R(kq) \ge \lambda\varphi(kq)\varphi^{\tau}(kq)$ as indicated in Equation (3), we have

$$\begin{aligned}&E\left[T(kq)|F_{kq-q}\right]\\ &\le (1-\lambda)T(kq-q) + E\left[\lambda\tilde{\theta}^{\tau}(kq-q)A\tilde{\theta}(kq-q)|F_{kq-q}\right]\\ &\quad + \lambda^2 E\left[\varphi^{\tau}(kq)R^{-1}(kq)\varphi(kq)|F_{kq-q}\right]\sigma_{\upsilon}^2 + E\left[w^{\tau}(kq)R(kq)w(kq)|F_{kq-q}\right] \qquad \text{(B.3)}\end{aligned}$$

Let $\beta = 1-\lambda$, and β satisfies $0 < \beta \le 1$. Simultaneously, we derive

$$R(kq) = \beta^k R_0 + \lambda\sum_{j=0}^{k-1}\beta^j\varphi(kq-jq)\varphi^{\tau}(kq-jq) + (1-\beta^k)A$$

From Equation (3) and A2, we can obtain

$$\frac{1}{pN}\sum_{i=0}^{pN-1}R(kq+iq) \le (1-\beta^k)\gamma I + \beta^k(R_0 - A) + A \qquad \text{(B.4a)}$$

$$\frac{1}{pN}\sum_{i=0}^{pN-1}R(kq+iq) \ge (1-\beta^k)\alpha I + \beta^{k+pN}(R_0 - A) + A \qquad \text{(B.4b)}$$

Based on A5, we know that $R_0 > A$. Hence, when $k > k_1$, the following equations hold:

$$(1-\beta^{k_1})\alpha I + A \le ER(kq) \le \gamma I + \beta^{k_1}(R_0 - A) + A \qquad \text{(B.5a)}$$

$$ER^{-1}(kq) \le \frac{1}{(1-\beta^{k_1})\alpha}I \qquad \text{(B.5b)}$$

$$\alpha I + A \leq \lim_{k\to\infty} E\left(R(kq)\right) \leq \gamma I + A \tag{B.5c}$$

On the other hand, we have

$$\lambda\tilde{\theta}^\tau(kq-q)A\tilde{\theta}(kq-q) \leq \lambda\lambda_{\max}(A)\frac{\tilde{\theta}^\tau(kq-q)R(kq-q)\tilde{\theta}(kq-q)}{\lambda_{\min}(R(kq-q))}$$

According to A5 and Equation (B.5), we have

$$E\left[\lambda\tilde{\theta}^\tau(kq-q)A\tilde{\theta}(kq-q)|F_{kq-q}\right] \leq \frac{\lambda\lambda_{\max}(A)}{(1-\beta^{k_1})\alpha+\lambda_{\min}(A)}T(kq-q) \tag{B.6}$$

and

$$\begin{aligned}-\lambda + \frac{\lambda\lambda_{\max}(A)}{(1-\beta^{k_1})\alpha+\lambda_{\min}(A)} &= \frac{-\lambda\lambda_{\min}(A)-\lambda(1-\beta^{k_1})\alpha+\lambda\lambda_{\max}(A)}{(1-\beta^{k_1})\alpha+\lambda_{\min}(A)} \\ &= -\frac{\lambda\left[(1-\beta^{k_1})\alpha-(\lambda_{\max}(A)-\lambda_{\min}(A))\right]}{(1-\beta^{k_1})\alpha+\lambda_{\min}(A)} \leq 0 \end{aligned} \tag{B.7}$$

According to A4 and Equations (B.5)-(B.7), Equation (B.3) satisfies

$$\begin{aligned}&E\left[T(kq)|F_{kq-q}\right] - T(kq-q) \\ &\leq -\frac{\lambda\left[(1-\beta^{k_1})\alpha-(\lambda_{\max}(A)-\lambda_{\min}(A))\right]}{(1-\beta^{k_1})\alpha+\lambda_{\min}(A)}T(kq-q) + \frac{\lambda^2\sigma_v^2M}{(1-\beta^{k_1})\alpha} \\ &\quad + \left[\gamma+\beta^{k_1}\lambda_{\max}(R_0-A)+\lambda_{\max}(A)\right]\sigma_w^2 = -b(kq)\end{aligned} \tag{B.8}$$

Let

$$\xi = \frac{\lambda^2\sigma_v^2M}{(1-\beta^{k_1})\alpha} + \left[\gamma+\beta^{k_1}\lambda_{\max}(R_0-A)+\lambda_{\max}(A)\right]\sigma_w^2$$

Notice that ξ is non-reduced. We consider the following collections:

$$R_k = \left\{\tilde{\theta}(kq)\middle|\tfrac{\lambda\left[(1-\beta^{k_1})\alpha-(\lambda_{\max}(A)-\lambda_{\min}(A))\right]}{(1-\beta^{k_1})\alpha+\lambda_{\min}(A)}T(kq-q) \leq \xi\right\}$$

$$N_\varepsilon(R_k) = \left\{\tilde{\theta}(kq)\middle|\tfrac{\lambda\left[(1-\beta^{k_1})\alpha-(\lambda_{\max}(A)-\lambda_{\min}(A))\right]}{(1-\beta^{k_1})\alpha+\lambda_{\min}(A)}T(kq-q) \leq \xi+\varepsilon,\ \varepsilon>0\right\}$$

Evidently, $R_k \subset N_\varepsilon(R_k)$. If $\tilde{\theta}(kq) \in N_\varepsilon^c(R_k)$, where $N_\varepsilon^c(R_k)$ represents the complementary set of $N_\varepsilon(R_k)$, then $E\left[T(kq)|F_{kq-q}\right] - T(kq-q) < 0$ from Equation (B.8) holds and $b(kq) \geq \varepsilon > 0$. As clearly shown in Lemma 3.2, $\tilde{\theta}(kq) \in N_\varepsilon(R_k)$ when k is sufficiently large. Given that ε is an arbitrary value, the conclusion that $\tilde{\theta}(kq) \in R_k$ always holds. In addition, when k_1 is sufficiently large, we can obtain

$$\begin{aligned}&\lim_{k_1\to\infty}\tilde{\theta}(kq) \in R_\infty \\ &= \left\{\begin{array}{l}\tilde{\theta}(kq)|\frac{\lambda[\alpha-(\lambda_{\max}(A)-\lambda_{\min}(A))]}{\alpha+\lambda_{\min}(A)}T(kq-q) \\ \qquad\qquad \leq \frac{\lambda^2\sigma_v^2M}{\alpha} + \left[(M\lambda^2-2\lambda)\gamma+\lambda_{\max}(A)\right]\sigma_w^2\end{array}\right\}\end{aligned}$$

Thus, the following equation holds:

$$\lim_{k\to\infty} E\left\|\tilde{\theta}(kq)\right\|^2 \leq \lim_{k\to\infty}\frac{T(kq)}{\lambda_{\min}(R(kq))} \leq \frac{\lambda^2\sigma_v^2M+\alpha\left(\gamma+\lambda_{\max}(A)\right)\sigma_w^2}{\alpha\lambda\left[\alpha-(\lambda_{\max}(A)-\lambda_{\min}(A))\right]} \tag{B.9}$$

The proof of Theorem 3.2 is completed.