Recursive Identification Algorithms Based
on Minimizing Estimation Error ⋆
Kueiming Lo ∗ Hidenori Kimura ∗∗ Donghai Li ∗∗∗ Wook-Hyun Kwon ∗∗∗∗

Abstract: Parameter selection for the criterion weighting matrix is concerned based on the information of both modifying the past estimation residuals and renewing the present estimation residual error. After minimizing the system estimation error, an optimal recursive algorithm is given. In this method the system data record can be used efficiently. The consistency of the new recursive algorithm is analyzed. Finally, some simulation examples are included to demonstrate the new method’s reliability.

Keywords: system identification, recursive algorithm, optimal recursive algorithm

1. INTRODUCTION

Since the measured input-output data is processed sequentially, adaptation schemes are adopted to re-identify the system online or in real-time operation and form the recursive identification algorithm. With rapid development of computer science and automatic control, this method has gained considerable attention and appeared much work, especially the celebrated RLS algorithm (For example, see Mersched and Sayed [2000], Lo and Zhang [2000], Hassibi and Kaliath [2001], Hellgren and Forssell [2001], and Ahn et al. [2004]. These recursive methods were based on the conventional prediction error (PE) criterion Ljung [1999].

Because of a real system complicated, it is difficult to get a precise prediction model. Smaller prediction errors does not mean smaller parameter estimation errors. In fact, if some complicated interferences occurred, a considerable identification error would arise from the RLS algorithms Lo and Kwon [2002], Lo and Kwon [2003], Yazdi et al. [2005], and Chan et al. [2006].

A general recursive algorithm for discrete systems is considered in this paper. First, a simplified recursive algorithm is proposed. There are many tuning parameters contained in this recursive algorithm. Some regulating techniques are established and the free tuning parameters are determined based on the system data record. Then, the identification principle is to construct an identification algorithm to estimate the system parameters. An optimal recursive algorithm is constructed by minimizing the parameter estimation error. This recursive algorithm is able to resist system noise, including color noise, biased noise, and noise for some unmodeled disturbance. Furthermore, the consistency of the optimal recursive algorithm is analyzed, and some simulation examples are included to demonstrate the new method’s reliability.

2. RECURSIVE IDENTIFICATION

Consider the SISO linear regression system:

\[ y_t = \varphi_t^T \theta + w_t \]  

where \( y_t \) is the system output, \( \varphi_t \in \mathbb{R}^n \) the regressor vector, \( \theta \) is the unknown system parameter and \( w_t \) the noise. The system identification often includes the following performance:

\[ J_t = [Y_t - \Phi_t \theta]^T Q_t [Y_t - \Phi_t \theta] \]  

where

\[ \Phi_t = (\varphi_1, \varphi_2, \ldots, \varphi_t)^T \]
\[ Y_t = (y_1, y_2, \ldots, y_t)^T \]

and \( \theta, \varphi_t \in \mathbb{R}^n \). The matrix \( Q_t \):

\[ Q_t = \begin{pmatrix} \alpha_t^{-1} \alpha_t^T & q_t \\ \alpha_t^T & q_t \end{pmatrix}, \quad \alpha_t \in \mathbb{R}^{t-1}, \quad t = 1, 2, 3, \ldots \]

is a \( t \times t \) symmetrical matrix. If the matrix \( \Phi_t^T Q_t \Phi_t \) is nonsingular, minimizing cost function (2), it is not difficult to get the optimal solution of parameter \( \theta \):

\[ \hat{\theta}_t = [\Phi_t^T Q_t \Phi_t]^{-1} \Phi_t^T Q_t Y_t, \quad t = 1, 2, 3, \ldots \]
Specifically, estimation (3) becomes a weighted LS algorithm if \( \alpha_t \) are taken as zeros. When \( \alpha_t \) are not zeros, the tuning parameter \( \hat{\theta}_t \) is complicated for every sample number \( t \). For this online calculation, computing \( \hat{\theta}_t \) is especially difficult when the sample number \( t \) increases. Let:

\[
P_t = \Phi_t Q_t \Phi_t^T
\]

\[
a_t = 1 + \varphi_t^T P_{t-1}^{-1} \alpha_t \tag{4}
\]

\[
\sigma_t = q_t - \alpha_t \Phi_{t-1} P_{t-1}^{-1} \varphi_t^T \tag{5}
\]

\[
b_t = a_t + \sigma_t \varphi_t^T P_{t-1}^{-1} \varphi_t \tag{6}
\]

The recursive algorithm of relation (3) was proposed as (Lo and Kimura [2003], Lo et al. [2006]):

\[
\begin{aligned}
\hat{\theta}_t &= \hat{\theta}_{t-1} + \frac{1}{a_t b_t} P_{t-1}^{-1} (a_t \Phi_{t-1}^T \alpha_t + \sigma_t \varphi_t) (y_t - \varphi_t^T \hat{\theta}_{t-1}) \\
& \quad + \frac{1}{a_t b_t} P_{t-1}^{-1} (a_t \Phi_{t-1}^T \alpha_t + \sigma_t \varphi_t) (Y_{t-1} - \Phi_{t-1} \hat{\theta}_{t-1}) \\
& \quad - \varphi_t^T P_{t-1}^{-1} \alpha_t \Phi_{t-1} \theta_t - \varphi_t^T P_{t-1}^{-1} \alpha_t \Phi_{t-1} \theta_{t-1} \\
P_t^{-1} &= P_{t-1}^{-1} - b_t^{-1} P_{t-1}^{-1} (\varphi_t \alpha_t^T \Phi_{t-1} + \Phi_{t-1} \alpha_t \varphi_t^T) P_{t-1}^{-1} \\
& \quad + \frac{1}{a_t b_t} P_{t-1}^{-1} (\varphi_t \alpha_t^T \Phi_{t-1} + \Phi_{t-1} \alpha_t \varphi_t^T) P_{t-1}^{-1} \Phi_{t-1} \\
& \quad - \sigma_t \varphi_t^T P_{t-1}^{-1} \varphi_t 
\end{aligned}
\tag{7}
\]

Algorithm (7) is composed of two decoupled complementary parts. One part renews the information of the current estimation residual; the other part modifies the estimation on past arithmetic errors. In this scheme it is easy to see that at time \( t \) there are \( t \) tuning parameters: \( q_t \) and \( \alpha_t \in \mathbb{R}^t \). Therefore, a common question is posed: how are variables chosen to guarantee that the algorithm is more reliable? If \( \alpha_t \) is chosen as: \( \alpha_t = 0, t = 1, 2, \ldots \) and \( q_t \) is a positive constant, from expressions (4)-(6), we have \( a_t = 1, \sigma_t = q_t, \) and

\[
b_t = 1 + q_t \varphi_t^T P_{t-1}^{-1} \varphi_t
\]

It follows that algorithm (7) is the same as the RLS algorithm, which only considers the present estimation residual in the modification part.

Due to a large number of free variables as well as the tuning parameters in algorithm (7), there is space to improve the algorithm performance. By minimizing the frequency-domain estimate, the recursive empirical frequency-domain optimal parameter (REFOP) estimate was proposed (Lo and Kimura [2003] and Lo et al. [2006]). In that algorithm, the tuning parameters could be generated by the input signal and the output signal of a given discrete system. Since it is derived from the frequency domain, the calculation seems somewhat complex. It could not make the choice of tuning parameters efficient, either. In this paper, we propose an optimal recursive algorithm (ORA) instead of the REFOP method. In the proposed algorithm, the tuning parameters are determined by minimizing the estimation error directly.

**Theorem 1.** With the previous notations, algorithm (7) can be expressed as:

\[
P_t^{-1} = P_{t-1}^{-1} - b_t^{-1} P_{t-1}^{-1} (\varphi_t \alpha_t^T \Phi_{t-1} + \Phi_{t-1} \alpha_t \varphi_t^T) P_{t-1}^{-1} \\
+ \frac{1}{a_t b_t} P_{t-1}^{-1} (\varphi_t \alpha_t^T \Phi_{t-1} + \Phi_{t-1} \alpha_t \varphi_t^T) P_{t-1}^{-1} \Phi_{t-1} \\
- \sigma_t \varphi_t^T P_{t-1}^{-1} \varphi_t
\]

\[
\hat{\theta}_t = \hat{\theta}_{t-1} + P_{t-1}^{-1} \left( \left[ \Phi_{t-1}^T \alpha_t + q_t \varphi_t \right] (y_t - \varphi_t^T \hat{\theta}_{t-1}) + \varphi_t \alpha_t^T (Y_{t-1} - \Phi_{t-1} \hat{\theta}_{t-1}) \right)
\tag{9}
\]

The proof of Theorem 1 is omitted.

Theorem 1 is another form of the recursive algorithm of estimation (3). It comes from algorithm (7); however, instead of the calculated priority order of \( \hat{\theta}_t \) and \( P_t^{-1} \) at time \( t \) in algorithm (7), the first calculation in Theorem 1 is that of \( P_t^{-1} \), and then the parameter \( \hat{\theta}_t \) is calculated by the result of \( P_t^{-1} \). Since of the complicated calculation in algorithm (7), it is difficult to obtain the following algorithm. Therefore, Theorem 1 is necessary for the next analysis.

### 3. OPTIMAL RECURSIVE ALGORITHM

**Lemma 1.** Suppose that matrix \( P_{t-1} \) is positive definite. Then matrix \( P_t \) is also positive definite if

\[
q_t > \alpha_t^T \Phi_{t-1}^{-1} \Phi_t^{-1} \alpha_t - (\varphi_t^T P_{t-1}^{-1} \varphi_t)^{-1} a_t^2
\tag{10}
\]

The proof of Lemma 1 is omitted.

If the initial matrix \( P_0 \) is positive definite and \( q_t, (t = 1, 2, 3, \ldots) \) satisfies the condition, Lemma 1 implies that the matrix \( P_t \) is also positive definite. Therefore, to optimize the recursive algorithm, the variable \( \alpha_t \) can be chosen by minimizing

\[
\tilde{\theta}_t^T P_t \tilde{\theta}_t, \text{ or } \tilde{\theta}_t^T P_t \tilde{\theta}_t, \text{ or } \tilde{\theta}_t^T P_t^2 \tilde{\theta}_t
\]

etc. Since

\[
\tilde{\theta}_t^T P_t^2 \tilde{\theta}_t = ||\tilde{\theta}_t^T P_t||^2
\]

it is easy to see that \( \tilde{\theta}_t = 0 \) if

\[
\tilde{\theta}_t^T P_t = 0
\]

Furthermore, since the function \( \tilde{\theta}_t^T P_t^2 \tilde{\theta}_t \) is more convenient to analyze some properties, it is chosen as the performance function in the following discussion.

**Theorem 2.** If the vector

\[
\alpha_t = (\alpha_t(1), \alpha_t(2), \ldots, \alpha_t(t-1))^T \in \mathbb{R}^{t-1}
\]

is chosen such that:

\[
(W_{t-1} \varphi_t^T + \Phi_{t-1} w_t) (\varphi_t W_{t-1}^T + \Phi_{t-1}^T w_t) \alpha_t = (W_{t-1} \varphi_t^T + \Phi_{t-1} w_t) (P_{t-1} \tilde{\theta}_{t-1} - q_t \varphi_t w_t)
\tag{11}
\]

then performance function

\[
J_t = \tilde{\theta}_t^T P_t \tilde{\theta}_t
\]

is the minimum, where:

\[
\tilde{\theta}_t = \theta - \tilde{\theta}_t, \quad W_t = (w_1, w_2, \ldots, w_t)^T
\]

The proof of Theorem 2 is omitted.
Remark 1. The number of the equations in relationship (12) is only $n$ since $\theta \in \mathbb{R}^n$. If $t$ is large enough, $n$ vectors $\varphi_{k1}, \ldots, \varphi_{kn}$ with $\varphi_{kn} \in \{\varphi_{1}, \varphi_{2}, \ldots, \varphi_{1}\}$ can be chosen such that matrix
\[
\Phi^T w_t + \varphi_{1}\mathbf{W}^T
\]
is nonsingular, where
\[
\Phi = (\varphi_{k1}, \ldots, \varphi_{kn})^T
\]
and
\[
\mathbf{W} = (w_{k1}, \ldots, w_{kn})^T
\]
are obtained. Hereafter, it is discussed how to estimate $\theta$. However, this result is based on the following requirements:

\[\alpha_{t}(l) = \begin{cases} 1 & l = k_i, i = 1, 2, \ldots, n \\ 0, & \text{others} \end{cases} \]

In a simple implementation, however, for a fix number $n$ we can revert a default positive value and use a switching technique for the matrix
\[
\mathbf{R}^T w_t + \varphi_{1}\mathbf{W}^T
\]
in case of singularity (see Example 1). Similar to the proof of Theorem 2, the condition:
\[
(\mathbf{W} \varphi_{1}^T + \Phi \varphi_{1} w_t^T + \Phi^T w_t \varphi_{1}) = (\mathbf{W} \varphi_{1}^T + \Phi \varphi_{1} (P_{t-1}\tilde{\theta}_{t-1} - q_t \varphi_{t} w_t))
\]
is equivalent to relationship (11). Since the nonsingular nature of the matrix
\[
\Phi^T w_t + \varphi_{1}\mathbf{W}^T
\]
the minimum of performance function $J_t$ can be achieved if the variable $\varphi_{1}$ is chosen as:
\[
\varphi_{1} = (\Phi^T \mathbf{W})^{-1} (P_{t-1}\tilde{\theta}_{t-1} - q_t \varphi_{t} w_t)
\]
(13)

Remark 2. From (12) it is easy to see that $\tilde{\theta}_t = 0$ if matrix $P_t$ is nonsingular and relation (13) is satisfied. This means that the real parameter $\theta$ can be obtained exactly at time $t$. In situations when symbols may not possibly be confused with each other, in the following discussion the symbols $\alpha_{t}$ and $\varphi_{t}$, $\tilde{\theta}_t$ and $\Phi$, and $W_{t}$ are not distinguished from each other.

However, this result is based on the following requirements: the signal $\{w_{k}\}_1^l$ and the difference $\tilde{\theta}_{t-1}$ are the estimation of $\theta_t$ and $w_t$ respectively. In practice it is impossible to get the precise information of signals $\{w_{k}\}_1^l$ and $\tilde{\theta}_{t-1}$ in system identification. Thus, $\varphi_{t}$ is only obtained by estimation:
\[
\alpha_{t} = \frac{1}{\varphi_{t}^T \mathbf{W}} - 1 (P_{t-1}\tilde{\theta}_{t-1} - q_t \varphi_{t} w_t)
\]
(14)

where $\tilde{\theta}_{t-1}$ and $\tilde{\theta}_{t-1}$ are the estimation of $\theta_{t-1}$ and $w_{t-1}$ respectively. At this time the selection of $k_1, k_2, \ldots, kn$ should satisfy the condition that the matrix
\[
\Phi^T \tilde{\theta}_{t-1} + \varphi_{1} \mathbf{W}^T
\]
is nonsingular. Hereafter, it is discussed how to estimate $\tilde{\theta}_{t-1}$. From system (1) the noise can be expressed as:
\[
w_t = y_t - \varphi_{t}^T \theta
\]
The most natural way to estimate noise is:
\[
\tilde{\theta}_{t-1} = y_t - \varphi_{t}^T \tilde{\theta}_{t-1}
\]
(15)

The ORA method is made up of relationships (8), (9), (14), and (15). The justification of such an estimation is verified by the next theorem.

Theorem 3. Let $q_t$ be chosen:
\[
q_t \in \left\{ \frac{2(|a_t | + (2 - \varepsilon) ||\varphi_{1} P_{t-1} \varphi_{t}^T \varphi_{t}^T P_{t-1} - a_t^2 ||^2}{1 - 2 \varepsilon a_t} + (1 - \varepsilon) \varphi_{1} P_{t-1} \varphi_{t}^T, \varepsilon \varphi_{t}^T P_{t-1} \varphi_{t}^T \right\}
\]
(16)

and the parameters $\tilde{\theta}_t$ be estimated by algorithms (8), (9), (14), and (15). Then we have
\[
||\tilde{\theta}_t|| \leq (1 - \varepsilon)||\tilde{\theta}_{t-1}||
\]
where $\varepsilon$ is a small positive number.

The proof of Theorem 3 is omitted.

Theorem 3 demonstrates that for an appropriate $q_t$, the estimation based on algorithms (8), (9), (14), and (15) is consistent. Since the algorithm contains relationship (15), the result of Theorem 3 also conforms to Theorem 3.

The proof of Theorem 3 is omitted.
Lemma 1 and Theorem 3 give the choice of domain for \( q_t \). Of course, this is a theoretic result for analyzing the performance of the new identification algorithm. In the following simulations, in fact, \( q_t \) is assigned a number from an interval \((0,1] \).

4. SIMULATIONS

Most of the early system identification work was based on the assumption that the interference system noise was either Gaussian, m.d.s., signals, or white noise. In engineering, these restrictions do not reflect reality since system noises are unknown and are very complicated. In this discussion there is no restriction on the system noise; that is, the disturbance may be non-Gaussian, non-m.d.s., or non-white noise. This performance is demonstrated by the following examples.

To illustrate the behavior of the optimal recursive algorithms, a simulation trial was conducted for comparison with the ordinary LS recursive algorithms, which was a special case of algorithms (8) and (9) with \( \alpha_t = 0 \), \( t = 1,2,3, \ldots \)

For a real system, the output \( \{y_t\}_t^N \) was generated by the system with a given input sequence \( \{u_t\}_t^N \). The experimental sample number \( N \) was 2000. Let

\[
\eta = (\sum_{t=1}^{N} w_t^2 / \sum_{k=1}^{N} y_k^2)^{1/2}
\]

be the noise-to-signal ratio, which expresses the extent of model signal disturbance. \( \theta \) was denoted as the real model parameter, while \( \hat{\theta}_{ORA} \) and \( \hat{\theta}_{LS} \) were the optimal recursive algorithm and the recursive least-squares (RLS) estimates.

Example 1. The discrete system was given as:

\[
y_t = \frac{b_1}{1 + a_1q^{-1}} w_{t-1} + w_t
\]  

(20)

The real system parameters were \( a_1 = 0.8 \) and \( b_1 = 3 \). The input signal \( \{u_t\}_t^N \) was generated by a sine generator. \( \{w_t\}_t^N \) was a stochastic disturbance with mean 1.55 and variance 0.52. It was a biased noise rather than a white noise. The output of the system was then generated by (20) with the noise-to-signal ratio \( \eta = 0.1475 \). The parameter was estimated according to the RLS method and the ORA1 method, which consisted of relationships (8), (9), (14), (15), and (18). The initial parameter \( \hat{\theta}_0 \) was the zero vector. At time \( t \) the matrix \( \mathbf{\tilde{\Phi}} \) was chosen as

\[
\mathbf{\tilde{\Phi}} = (\varphi_{t-2}, \varphi_{t-1})^T
\]

To avoid the singularity of the matrix

\[
\varphi_t \mathbf{W}^T + \mathbf{\tilde{\Phi}}^T \mathbf{\tilde{\omega}}_t
\]

in (14), the \( \mathbf{\tilde{\omega}}_t \) was chosen as the zero vector if

\[
\left| \det(\varphi_t \mathbf{W}^T + \mathbf{\tilde{\Phi}}^T \mathbf{\tilde{\omega}}_t) \right| < 0.1
\]

In fact, only six switches were encountered in the full simulation.

The optimal recursive algorithm can be intuitively compared with the RLS method as shown in Figure 1:

\[
\eta_{ORA1} = \begin{pmatrix} 0.7916 \pm 0.0025 \\ 3.6084 \pm 0.0206 \end{pmatrix}, \quad \eta_{ORA2} = \begin{pmatrix} 0.4100 \pm 0.0194 \\ 2.3531 \pm 0.0307 \end{pmatrix}
\]

where \( \eta_{ORA1} \) denotes the average estimate from the 100th ORA value to the 2000th ORA value. \( \eta_{ORA2} \) denotes the average estimate from the value of the 100th LS estimate to the 2000th LS estimate. The calculation error is defined by standard deviation.

Example 2. The discrete system is given as:

\[
y_t = \frac{b_{t-1}}{1 + a_1 q^{-1} + a_2 q^{-2}} + w_t
\]

(21)

where

\[
w_t = \frac{w_{1t}}{1 + q^{-1} + 0.2q^{-2}} + w_{2t}
\]

The real system parameters were \( a_1 = -1.7, a_2 = 0.8 \). The input signal \( \{u_t\}_t^N \) was generated by a pulse generator. \( \{w_{1t}\}_t^N \) was an approximate white noise and \( \{w_{2t}\}_t^N \) was a sawtooth signal. The noise \( \{w_{1t}\}_t^N \) consisted of these two signals with mean 0.0301 and variance 2.0466. As in example 1, this noise was not a white noise, either. The output of the system was then generated by (21) with a noise-to-signal ratio \( \eta = 0.2237 \). The parameter was estimated according to the RLS method and the ORA2 method, which consisted of relationships (8), (9), (14), (15), and (19) with \( m = 5 \). From (19) we can see that the desired estimation \( \hat{\theta}_{t-1} \) would depend on the choice of initial parameter \( \hat{\theta}_0 \). In order to get better initial values, the recursive empirical frequency-domain optimal estimate was used in the first 200 steps, where \( \mathbf{\tilde{\alpha}}_t \) was determined by (8) and \( v(t) = y_t \) (Lo and Kimura 2003). At time \( t \) matrix \( \mathbf{\tilde{\Phi}} \) was chosen as

\[
\mathbf{\tilde{\Phi}} = (\varphi_{t-2}, \varphi_{t-1})^T
\]

too. To avoid the singularity of the matrix

\[
\varphi_t \mathbf{W}^T + \mathbf{\tilde{\Phi}}^T \mathbf{\tilde{\omega}}_t
\]
in (14), $\theta_t$ was chosen as the zero vector if
\[
|\text{det}(\varphi_t W^T + \Phi^T \hat{w}_t)| < 0.1
\]
However, no switch occurred in the full simulation. Figure 2 shows the simulation result.

Due to the interference of complicated and biased noise, a large error occurred using the ordinary RLS algorithm. Due to the interference of complicated and biased noise, Figure 2 shows that the estimation of the ORA2 method was more precise than that of the RLS method. The calculated values were given by the following:
\[
\overline{\theta}_{ORA2} = \begin{pmatrix}
-1.6969 & \pm 0.0038 \\
0.8000 & \pm 0.0047
\end{pmatrix}
\]
\[
\overline{\theta}_{LS} = \begin{pmatrix}
-0.9769 & \pm 0.0383 \\
0.0860 & \pm 0.0387
\end{pmatrix}
\]
where $\overline{\theta}_{ORA2}$ denotes the average estimate from the 100th ORA value to the 2000th ORA value. $\overline{\theta}_{LS}$ denotes the average estimate from the value of the 100th LS estimate to the 2000th LS estimate.

5. CONCLUSIONS

This paper presents a recursive algorithm for discrete systems, composed of two decoupled complementary parts. There are many tuning parameters contained in this recursive algorithm. They are included and constructed in a time-varying performance criterion and in the recursive algorithm. Minimizing the parameter estimation error, an ORA method was established. Compared with the previous recursive algorithms, in this method the system data record can be used efficiently, and computational time is less than that of the REFOP estimate (Lo and Kimura [2003]). Furthermore, the viability and consistency of the optimal recursive algorithm were analyzed. Since there was not any restriction on the system disturbance noise, theoretic analysis and simulation results indicated that the new algorithms have the advantage of being anti-interference, which includes protection against color noise, biased noise, and noise for some unmodeled systems.

The purpose of our research is to extend and adapt system identification to be efficiently used in complicated engineering applications. As a special case of algorithms (8), (9), and (22), the LS algorithm should be the simplest choice because
\[
\alpha_t = 0, \quad t = 1, 2, 3, \ldots
\]
REFERENCES


