NEW RECURSIVE LEAST SQUARE ALGORITHMS WITHOUT USING THE INITIAL INFORMATION

Zhonghua Quan * **Soohee Han** ** **Wook Hyun Kwon** *

* *School of Electrical Engineering & Computer Science, Seoul National University, Seoul, 151-742, Korea*
** *Engineering Research Center for Advanced Control & Instrumentation, Seoul National University, Seoul, 151-742, Korea*

Abstract: In this paper, new types of recursive least square (RLS) algorithms, without using the initial information of a parameter or a state to be estimated, are proposed. The proposed RLS algorithm is first obtained for a generic linear model and is then extended to a state estimator for a stochastic state-space model. Compared with the existing algorithms, the proposed RLS algorithms are simpler and more numerically stable. It is shown, by simulation studies, that the proposed RLS algorithms have better numerical stability for digital computation than existing algorithms.

Keywords: Estimation, initial information, Recursive Least-Squares, state estimator, numerical stability.

1. INTRODUCTION

Estimation problems deal with the determination of some unknown variables, or parameters, that cannot be measured directly. The least square (LS) approach, tracing its origins back to Gauss in the nineteenth century, is certainly widely used in many estimation problems. In particular, the recursive least square (RLS) algorithm is known to be more practical and efficient in computation than the LS algorithm of the batch form.

The RLS algorithm provides good performance for models with the accurate initial information for the parameter or state to be estimated. If the assumed initial information is not correct, however, the performance deteriorates. In many cases, it may be difficult to obtain the correct initial information because it may be costly and require many experiences. Therefore, it is necessary to find RLS algorithms that do not require the initial information when it is impossible to obtain the correct initial information.

For generic linear models, some attempts have been made to find an RLS algorithm without the requirement of the initial information. The QR decomposition-based RLS algorithm introduced in (Gentlemen and Kung, 1981) and (Haykin, 1996) does not need the initial conditions, however, it requires that the data matrix, and therefore the square root of the correlation matrix of the input data, is of full column rank for the feasibility. In (Danyang and Xuanhuang, 1994), an LS algorithm is proposed using a recursive form of the pseudo-inverse of a large matrix presented in (Cline, 1964), (Greville, 1960) and (Greville, 1959), which do not need the full column rank condition. However, it is computationally demanding and many pseudo-inverse operations are required, which may result in numerical instability or even divergence in practice because the pseudo-inverse can be ill-conditioned with respect to small rank-changing perturbations. In (Jie Zhou and You, 2001) and (Jie Zhou and You, 2002), an RLS algorithm based on Greville's formula is proposed by Jie Zhou *et al.*, where the problem of numerical instability also occurs as in (Danyang and Xuanhuang, 1994). To achieve numerical stability, it is necessary to use the pseudo-inverse as little as possible, which is the motivation for this work.

For state-space models, the Kalman estimator (Kalman, 1960) is the most widely used state estimator. It is dependent on the initial information of a state. A state estimator, without the requirement of any a priori information about the initial state, was presented in (Danyang and Xuanhuang, 1994) for state-space models. This algorithm encounters the problem of numerical instability due to too many pseudo inversions as in the RLS algorithm for generic linear models.

In this paper, a new type of RLS algorithm for the generic linear model is developed without requiring the initial information of the parameter to be estimated. In addition, a new state estimator without requiring any initial information of the state is proposed for state-space models with system and measurement noises. The proposed RLS algorithms are numerically stable because the pseudo-inverse matrix is used only once, which make it different from other approaches where several pseudo-inverse operations are required. By simulation, it is shown that the proposed RLS algorithms have greater numerical stability in digital computation than existing algorithms. In addition, they are simpler than existing algorithms and can be readily applied to real estimation problems.

This paper is organized as follows. In Section II, a new type of RLS algorithm is presented for a simple generic linear model and this is extended to a state estimator without using any a priori initial information of a state for the stochastic state-space model in Section III. In Section IV, simulation results are given to validate the RLS algorithms proposed in this paper. Finally, conclusions are given in Section V.

2. PSEUDO-INVERSE BASED RLS ALGORITHM

Consider a following generic linear model:

$$z_{k+1} = h_{k+1}\theta + \epsilon_{k+1}, \qquad k = 0, 1, 2, ... \quad (1)$$

where $\theta \in \Re^r$ is an unknown random vector to be estimated, and z_k, h_k and ϵ_k are a measurement, a known constant row vector and an unknown disturbance, respectively. Then, $\hat{\theta}$ is the estimated value of θ that minimizes the following weighted Euclidean norm

$$\min_{\theta} \left[\theta^* \Pi_0^{-1} \theta + \| Z_{k+1} - H_{k+1}\theta \|^2_{L_{k+1}^{-1}} \right], \quad (2)$$

where

$$H_{k+1} = \left[h_1^T \cdots h_{k+1}^T \right]^T = \left[H_k{}^T \;\; h_{k+1}^T \right]^T \quad (3)$$

$$Z_{k+1} = \left[z_1^T \cdots z_{k+1}^T \right]^T = \left[Z_k{}^T \;\; z_{k+1}^T \right]^T \quad (4)$$

$$L_{k+1} = diag\{ l_1 \cdots l_{k+1} \} = diag\{ L_k \; l_{k+1} \} \quad (5)$$

and L_k is a weight matrix. The solution of (2) can be computed using the following RLS algorithm:

$$\hat{\theta}_{k+1} = \hat{\theta}_k + K_{k+1}(z_{k+1} - h_{k+1}\hat{\theta}_k) \quad (6)$$

$$K_{k+1} = P_k h_{k+1}^T / (l_{k+1} + h_{k+1} P_k h_{k+1}^T) \quad (7)$$

$$\begin{aligned} P_{k+1} &\stackrel{def}{=} (\Pi_0^{-1} + H_{k+1}^T L_{k+1}^{-1} H_{k+1})^{-1} \\ &= P_k - \frac{P_k h_{k+1}^T h_{k+1} P_k}{l_{k+1} + h_{k+1}^T P_k h_{k+1}} \end{aligned} \quad (8)$$

with $\hat{\theta}_0 = 0$ and $P_0 = \Pi_0$.

When H_k in (8) is not of full column rank, i.e., the matrix P_k is not defined, then (8) is not valid. Therefore, pseudo-inverse-based RLS algorithms were proposed in (Danyang and Xuanhuang, 1994), (Jie Zhou and You, 2001) and (Jie Zhou and You, 2002). These are derived from the recursive form of the following formula:

$$\hat{\theta}_k = H_k^{'+} Z_k^{'}, \quad (9)$$

where $H_k^{'} = L_k^{-\frac{1}{2}} H_k$, $Z_k^{'} = L_k^{-\frac{1}{2}} Z_k$, and H^+ is defined as the unique matrix such that $HH^+H = H$, $H^+HH^+ = H^+$, $(H^+H)^T = H^+H$, and $(HH^+)^T = HH^+$ are satisfied.

The RLS algorithms in (Danyang and Xuanhuang, 1994), (Jie Zhou and You, 2001) and (Jie Zhou and You, 2002), however, use too many pseudo-inverses which may result in instability in the numerical computations with finite machine precision because the pseudo-inverse is ill-conditioned with respect to rank-changing perturbations. For example,

$$\begin{bmatrix} 1+\varepsilon & -1 \\ 2 & 2(\varepsilon - 1) \end{bmatrix}^+ = \begin{cases} \dfrac{1}{\varepsilon^2} \begin{pmatrix} -1 & \frac{1}{2} \\ -1 & \frac{1}{2} \end{pmatrix} + \dfrac{1}{\varepsilon} \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{pmatrix} & \text{if } \varepsilon \neq 0 \\ \dfrac{1}{10} \begin{pmatrix} 1 & 2 \\ -1 & -2 \end{pmatrix} & \text{if } \varepsilon = 0 \end{cases} \quad (10)$$

It can be seen that as $\varepsilon \to 0$, the matrix in the case of $\varepsilon \neq 0$ in (10) diverges. However, for fixed $\varepsilon = 0$, a constant matrix with finite elements is obtained. Therefore, it is necessary to use the pseudo-inverse as little as possible for numerical stability.

Now let us derive a new type of RLS algorithm. Because H^+ can be written as $H^+ = (H^T H)^+ H^T$(Rao and Mitra, 1971), the $\hat{\theta}_k$ in (9) can be written as:

$$\hat{\theta}_k = (H_k^T L_k^{-1} H_k)^+ H_k^T L_k^{-1} Z_k. \quad (11)$$

From (11), the new RLS algorithm can be derived, which is introduced in the following theorem.

Theorem 1. *The new RLS algorithm can be described as:*

$$\begin{aligned} \hat{\theta}_{k+1} &= \Sigma_{k+1}^+ \Gamma_{k+1} \\ \Sigma_{k+1}^+ &= \Sigma_k^+ - \Sigma_k^+ h_{k+1}^T \Upsilon_{k+1}^{-1} h_{k+1} \Sigma_k^+ \\ \Gamma_{k+1} &= \Gamma_k + h_{k+1}^T l_{k+1}^{-1} z_{k+1} \\ \Upsilon_{k+1} &= l_{k+1} + h_{k+1} \Sigma_k^+ h_{k+1}^T \end{aligned} \quad (12)$$

where $\Gamma_0 = 0$ *and* $\Sigma_0^+ = 0$.

Proof. Let $\Sigma_k = H_k^T L_k^{-1} H_k$ and $\Gamma_k = H_k^T L_k^{-1} Z_k$, then from (11), we have

$$\hat{\theta}_k = \Sigma_k^+ \Gamma_k, \quad (13)$$

and the recursive forms of the Σ_k^+ and Γ_k can be easily obtained from (3) and (4), respectively. This completes the proof. □

Remark 1. *Due to the use of the pseudo-inverse in (12), there is no need to guarantee the non-singularity of the matrix* Σ_k. *Furthermore, the proposed RLS algorithm is more stable than the existing algorithms because the matrix pseudo-inverse is used only once and it is not calculated recursively. If the matrix* Σ_k *is invertible, then* θ_k *can be obtained from* $\hat{\theta}_k = \Sigma_k^{-1} \Gamma_k$.

3. STATE ESTIMATOR WITHOUT THE REQUIREMENT OF INITIAL INFORMATION

Now let us investigate the problem on the state estimation in the state-space model. Consider a linear discrete-time state-space signal model:

$$\begin{aligned} x_{k+1} &= A_k x_k + B_k u_k + G_k w_k \\ y_k &= C_k x_k + v_k, \end{aligned} \quad (14)$$

where $x_k \in \Re^n$, $u_k \in \Re^l$, $y_k \in \Re^q$ are the state, the input, and measurement, respectively. The matrix A_k is assumed to be non-singular, which is not a strong condition because it is obtained by the discretization of a continuous-time state equation and is the exponential function of the system matrix in the continuous-time state equation.

The initial state x_0 is a random variable with a mean m_0 and a covariance P_0 . The system noise $w_k \in \Re^p$ and the measurement noise $v_k \in \Re^q$ are zero-mean white Gaussian and mutually uncorrelated. The covariances of w_k and v_k are denoted by Q and R, respectively, which are assumed to be positive definite matrices. These noises are uncorrelated with the initial state x_0.

To avoid the numerical instability problems described above, a new state estimator without using the initial information of the initial states is proposed.

For the state-space model represented by (14), the measurements over the time interval $[0, k]$ can be expressed in terms of the state x_k at the current time k as:

$$Y_{k-1} = \mathbb{C}_k x_k + \mathbb{B}_k U_{k-1} + \mathbb{G}_k W_{k-1} + V_{k-1} \quad (15)$$

where

$$Y_{k-1} \triangleq [y_0^T \ \cdots \ y_{k-1}^T]^T \quad (16)$$

$$U_{k-1} \triangleq [u_0^T \ \cdots \ u_{k-1}^T]^T \quad (17)$$

$$W_{k-1} \triangleq [w_0^T \ \cdots \ w_{k-1}^T]^T \quad (18)$$

$$V_{k-1} \triangleq [v_0^T \ \cdots \ v_{k-1}^T]^T \quad (19)$$

and $\mathbb{C}_k$, $\mathbb{B}_k$, $\mathbb{G}_k$ can be obtained from

$$\mathbb{C}_i \triangleq \begin{bmatrix} \mathbb{C}_{i-1} \\ C_{i-1} \end{bmatrix} A_{i-1}^{-1}, \quad 2 \le i \le k \quad (20)$$

$$\mathbb{B}_i \triangleq \begin{bmatrix} \mathbb{B}_{i-1} & -\mathbb{C}_{i-1} A_{i-1}^{-1} B_{i-1} \\ 0 & -C_{i-1} A_{i-1}^{-1} B_{i-1} \end{bmatrix} \quad (21)$$

$$\mathbb{G}_i \triangleq \begin{bmatrix} \mathbb{G}_{i-1} & -\mathbb{C}_{i-1} A_{i-1}^{-1} G_{i-1} \\ 0 & -C_{i-1} A_{i-1}^{-1} G_{i-1} \end{bmatrix}. \quad (22)$$

$\mathbb{C}_1$, $\mathbb{B}_1$, and $\mathbb{G}_1$ are defined as

$$\mathbb{C}_1 \triangleq C_0 A_0^{-1}, \quad \mathbb{B}_1 \triangleq -C_0 A_0^{-1} B_0, \quad \mathbb{G}_1 \triangleq -C_0 A_0^{-1} G_0.$$

The noise term $\mathbb{G}_k W_{k-1} + V_{k-1}$ in (15) can be shown to be zero-mean with covariance Ξ_k given by

$$\begin{aligned} \Xi_i &\triangleq \mathbb{G}_i \left[\mathrm{diag}(\overbrace{Q \cdots Q}^{i})\right] \mathbb{G}_i^T + \left[\mathrm{diag}(\overbrace{R \cdots R}^{i})\right] \quad (23) \\ &= \begin{bmatrix} \Xi_{i-1} & 0 \\ 0 & R \end{bmatrix} + \begin{bmatrix} \mathbb{C}_{i-1} \\ C_{i-1} \end{bmatrix} A_{i-1}^{-1} G_{i-1} Q G_{i-1}^T A_{i-1}^{-T} \begin{bmatrix} \mathbb{C}_{i-1} \\ C_{i-1} \end{bmatrix}^T \end{aligned}$$

The equation (15) can be rewritten as

$$[Y_{k-1} - \mathbb{B}_k U_{k-1}] = \mathbb{C}_k x_k + [\mathbb{G}_k W_{k-1} + V_{k-1}] \quad (24)$$

Note that (24) can be written more compactly in the form (2) with the identifications

$$\begin{aligned} &Z_k \longleftarrow \{Y_{k-1} - \mathbb{B}_k U_{k-1}\}, \quad H_k \longleftarrow \mathbb{C}_k, \\ &\theta_k \longleftarrow x_k, \quad \epsilon_k \longleftarrow \mathbb{G}_k W_{k-1} + V_{k-1}, \quad L_k \longleftarrow \Xi_k. \end{aligned}$$

Therefore, $\hat{x}_k$ can be obtained by

$$\hat{x}_k = \Omega_k^+ \eta_k \tag{25}$$

in the form (12) with the identifications

$$\Sigma_k \longleftarrow \Omega_k, \quad \Gamma_k \longleftarrow \eta_k.$$

where

$$\Omega_k \triangleq \mathbb{C}_k^T \Xi_k^{-1} \mathbb{C}_k$$
$$\eta_k \triangleq \mathbb{C}_k^T \Xi_k^{-1} (Y_{k-1} - \mathbb{B}_k U_{k-1}) \tag{26}$$

The above results in this section can be summarized by the following theorem.

Theorem 2. *In the singular region $0 \le k \le n-1$, $\hat{x}_k$ that minimizes the following weighted Euclidean norm*

$$\min_{\hat{x}_k} \|Y_{k-1} - \mathbb{B}_k U_{k-1} - \mathbb{C}_k \hat{x}_k\|^2_{\Xi_k^{-1}} \tag{27}$$

can be written as

$$\hat{x}_k = \Omega_k^+ \eta_k. \tag{28}$$

By tedious, but trivial, computation using (20) and (23) as in (WookHyun Kwon and Han, 2002), the following recursive forms of Ω_k^+ and η_k can be obtained, respectively:

$$\begin{aligned}
&\Omega_{k+1}^+ \\
&= (\mathbb{C}_{k+1}^T \Xi_{k+1}^{-1} \mathbb{C}_{k+1})^+ \\
&= \{A_k^{-T} \begin{bmatrix} \mathbb{C}_k \\ C_k \end{bmatrix}^T \left(\begin{bmatrix} \Xi_k & 0 \\ 0 & R \end{bmatrix} + \begin{bmatrix} \mathbb{C}_k \\ C_k \end{bmatrix} A_k^{-1} G_k Q G_k^T \right. \\
&\times A_k^{-T} \left. \begin{bmatrix} \mathbb{C}_k \\ C_k \end{bmatrix}^T \right)^{-1} \begin{bmatrix} \mathbb{C}_k \\ C_k \end{bmatrix} A_k^{-1} \}^+ \\
&= \{[I + A_k^{-T}(\Omega_k + C_k^T R^{-1} C_k) A_k^{-1} G_k Q G_k^T]^{-1} A_k^{-T} \\
&\times (\Omega_k + C_k^T R^{-1} C_k) A_k^{-1}\}^+ \\
&= A_k(\Omega_k^+ - \Omega_k^+ C_k^T \Pi_k^{-1} C_k \Omega_k^+) A_k^T [I + A_k^{-T}(\Omega_k^+ \\
&- \Omega_k^+ C_k^T \Pi_k^{-1} C_k \Omega_k^+)^{-1} A_k^{-1} G_k Q G_k^T]
\end{aligned} \tag{29}$$

and

$$\begin{aligned}
&\eta_{k+1} \\
&= \mathbb{C}_{k+1}^T \Xi_{k+1}^{-1} (Y_k - \mathbb{B}_{k+1} U_k) \\
&= [I + A_k^{-T}(\Omega_k + C_k^T R^{-1} C_k) A_k^{-1} G_k Q G_k^T]^{-1} A_k^{-T} \\
&\times [\eta_k + C_k^T R^{-1} y_k + (\Omega_k + C_k^T R^{-1} C_k) A_k^{-1} B_k u_k] \\
&= [I + A_k^{-T}(\Omega_k^+ - \Omega_k^+ C_k^T \Pi_k^{-1} C_k \Omega_k^+)^{-1} A_k^{-1} G_k Q \\
&\times G_k^T]^{-1} A_k^{-T} [\eta_k + C_k^T R^{-1} y_k + (\Omega_k^+ - \Omega_k^+ C_k^T \Pi_k^{-1} \\
&\times C_k \Omega_k^+)^{-1} A_k^{-1} B_k u_k].
\end{aligned} \tag{30}$$

with $\Omega_0^+ = 0$ and $\eta_0 = 0$, where $\Pi_k = R + C_k \Omega_k^+ C_k^T$. For $k \ge n$, Ω_k^+ is equal to Ω_k^{-1}.

Remark 2. *Note that no a priori information of the initial states is required in this algorithm.*

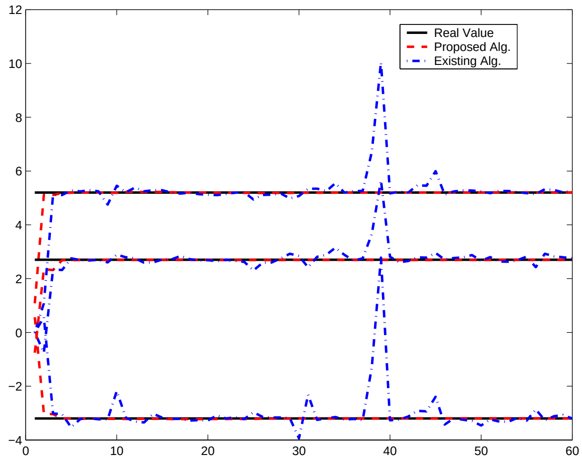

Fig. 1. Simulation for RLS

4. SIMULATION RESULTS

To demonstrate the validity of the proposed algorithms, numerical examples are presented from simulation studies. Measurement data used in the simulation for the proposed RLS algorithm are generated by the following linear model:

$$Y(k) = \begin{bmatrix} 5.2 & 2.7 & -3.2 \end{bmatrix} \begin{bmatrix} u_1(k) \\ u_2(k) \\ u_3(k) \end{bmatrix} + e(k) \tag{31}$$

where $e(k)$ is a zero mean white noise with a standard deviation of 0.5. Fig. 1 shows that the existing algorithms result in oscillation because of the product of an infinitesimal by infinity, which may produce NaN (Not a Number) in numerical computation. On the other hand, the parameters can be estimated more correctly using the proposed RLS algorithm, without oscillation, which means that the proposed RLS algorithm is more stable than the existing algorithms described above.

To compare the proposed estimator with the existing estimators, consider the system (14) with the system matrices

$$A = \begin{bmatrix} 0.9305 & 0 & 0.1107 \\ 0.0077 & 0.9802 & -0.0173 \\ 0.0142 & 0 & 0.8953 \end{bmatrix},$$
$$G = \begin{bmatrix} 0 & 1 & 2 \\ 0 & 2 & 3 \\ 0 & 3 & 9 \end{bmatrix}, \quad C = \begin{bmatrix} 1 + \sin k \\ 1 - 2\cos k \\ 1 + 4\cos k \end{bmatrix}^T \tag{32}$$

and zero input.

From Fig.2, it can be seen that the existing algorithm results in oscillation for the reasons described above. On the other hand, it can be observed that the proposed estimator works well without oscillation, which occurred in the results from previous methods. From this viewpoint, it may be said that the approach proposed is more

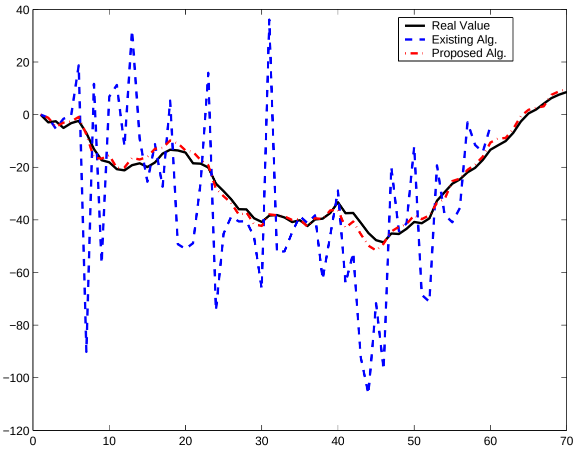

Fig. 2. Simulation for state estimator

stable than the existing approach. Furthermore, the estimation results for the state can track the real state from the start of the state estimation.

5. CONCLUSION

In this paper, new RLS algorithms that do not use the initial information are proposed for both generic linear models and state-space models. In the proposed RLS algorithm, the pseudo-inverse operation is taken only once, which is advantageous for numerical implementation. Simulation showed that the proposed algorithm is more efficient, numerically stable and accurate than previous methods. In addition, the proposed RLS algorithms are simpler than the existing algorithms. Therefore, they can be easily applied to many real estimation problems.

REFERENCES

Cline, Randall E. (1964). Representations for the generalized inverse of a partitioned matrix. *J.Soc.Indust.Appl.Math* **12**, 588–600.

Danyang, Liu and Liu Xuanhuang (1994). Optimal state estimation without the requirement of a priori statistics information of the initial state. *IEEE Trans. Automatic Control* **39**, 2087–2091.

Gentlemen, W. M. and H. T. Kung (1981). Matrix triangularization by systolic arrays. *Proc. SPIE* **298**, 298–303.

Greville, T.N.E. (1959). The pseudoinverse of a rectangular or singular matrix and its application to the solution of systems of linear equations. *SIAM Review* **1**, 38–43.

Greville, T.N.E. (1960). Some applications of the pseudoinverse of a matrix. *SIAM Review* **2**, 15–22.

Haykin, Simon (1996). *Adaptive filter theory.* Prentice-Hall. New Jersey.

Jie Zhou, X. Rong Li and Zhisheng You (2001). Exactly initialized recursive least squares. *Proceedings of the 40th IEEE Conference on Decision and Control* pp. 3318–3323.

Jie Zhou, Yunmin Zhu, X. Rong Li and Zhisheng You (2002). Variants of the greville formula with applications to exact recursive least squares. *SIAM J. Matrix ANAL. APPL.* **24**, 150–164.

Kalman, R.E. (1960). A new approach to linear filtering and prediction problems. *Transactions of the ASME-Journal of Basic Engineering* **82**, 35–46.

Rao, C. Radhakrishna and Sujit Kumar Mitra (1971). A new approach to linear filtering and prediction problems. *John Wiley & Sons, Inc.*

WookHyun Kwon, PyungSoo Kim and SooHee Han (2002). A receding horizon unbiased fir filter for discrete-time state space models. *Automatica* **38**, 545–551.