On the Uniqueness of Prediction Error Models for Systems with Noisy Input–Output Data*

PETRE STOICAI† and ARYE NEHORAI‡§

Key Words—System identification; parameter estimation; global identifiability; noisy systems; prediction error methods.

Abstract—This paper addresses the uniqueness problem of the prediction error (PE) identification for a class of linear systems with noisy input and output data. Necessary and sufficient conditions are derived for the corresponding PE loss function to have (asymptotically) a unique global minimum. The results indicate that a PE algorithm may give very bad parameter estimates for systems not satisfying these conditions. Such a possibility is illustrated by a numerical example. While the PE method is used as a vehicle for illustration, the derived conditions for global uniqueness (or identifiability) apply to any consistent estimation method based on second-order data.

1. Introduction and preliminaries

Consider a discrete time linear system characterized by

\[ w(t) = \frac{B(q^{-1})}{A(q^{-1})} x(t) \quad t = 1, 2, \ldots \]  \hspace{1cm} (1.1a)

where \( w(t) \) is the system output at time instant \( t \), \( x(t) \) is the input, and \( A(q^{-1}) \) and \( B(q^{-1}) \) are polynomials in the unit delay operator \( q^{-1} \) \([q^{-1}w(t) = w(t-1)]\):

\[ A(q^{-1}) = 1 + a_1 q^{-1} + \cdots + a_n q^{-n} \quad a_n \neq 0 \]
\[ B(q^{-1}) = b_0 + b_1 q^{-1} + \cdots + b_m q^{-m} \quad b_0 \neq 0, b_m \neq 0. \]  \hspace{1cm} (1.1b)

It is assumed that \( A(z) \neq 0 \) for \( |z| < 1 \), and that the polynomials \( A(z) \) and \( B(z) \) are coprime. Thus the representation (1.1) is stable with minimal orders \((n, m)\). It is further assumed that the (noise-free) input \( x(t) \) can be modeled as an autoregressive moving-average (ARMA) process,

\[ x(t) = \frac{G(q^{-1})}{H(q^{-1})} \varepsilon(t) \]  \hspace{1cm} (1.2a)

where \( \varepsilon(t) \) is a zero mean white noise of variance \( \lambda_0^2 \), and

\[ G(q^{-1}) = 1 + g_1 q^{-1} + \cdots + g_n q^{-n} \quad g_n \neq 0 \]
\[ H(q^{-1}) = 1 + h_1 q^{-1} + \cdots + h_m q^{-m} \quad h_m \neq 0. \]  \hspace{1cm} (1.2b)

Without loss of generality, it can be assumed that \( G(q^{-1}) \) and \( H(q^{-1}) \) are coprime polynomials, and that \( G(z) \neq 0, H(z) \neq 0, \) for \( |z| < 1 \) [see the spectral factorization theorem, e.g., in Åström (1970)]. Thus the ARMA representation (1.2) is stable and invertible with minimal orders \((n, m)\).

Let \( y(t) \) and \( u(t) \) be the noise-corrupted measurements of \( w(t) \) and \( x(t) \), respectively, i.e.

\[ y(t) = w(t) + e_y(t) \]
\[ u(t) = x(t) + e_x(t) \]  \hspace{1cm} (1.3)

where it is assumed that \( e_y(t) \) and \( e_x(t) \) are zero mean white noises of variances \( \lambda_y^2 \) and \( \lambda_x^2 \), respectively, independent of each other and of \( \varepsilon(t) \).

The unknown true parameter vector

\[ \theta = [\lambda_y^2, \lambda_x^2, \lambda_0^2, a_1, \ldots, a_n, b_0, \ldots, b_m]^T \]
\[ g_1, \ldots, g_n, h_1, \ldots, h_m]^T \]  \hspace{1cm} (1.4)

is to be estimated from \( N \) pairs of noisy input/output measurements \([u(1), y(1), \ldots, u(N), y(N)]\). The orders \( n, m \) etc. are assumed to be known.

It is worth noting here that the simplifying assumptions of known orders, white and independent noises are used in order not to conceal the main points of the subsequent analysis. A similar but more involved analysis would be required if these assumptions were relaxed.

Observe from (1.1)–(1.3) that the bivariate process \([y(t), u(t)]^T\) has a rational spectral density function. Thus, from the second order properties standpoint this process can be represented as

\[ \begin{bmatrix} x(t) \\ u(t) \end{bmatrix} = S_{gu}^{-1}(\theta) e(t, \theta) \]  \hspace{1cm} (1.5)

where \( E[e(t, \theta) e^T(t, \theta)] = S_{\theta}(\theta) \delta(t, \theta) \) is the Kronecker delta; \( S_{gu}^{-1}(\theta) \) is a \( 2 \times 2 \) matrix whose entries are rational functions of \( q^{-1} \), with \( S_{\theta}(\theta) = I \) and both \( S_{gu}^{-1}(\theta) \) and \( S_{\theta}(\theta) \) are asymptotically stable [see, e.g., Anderson and Moore (1979)]. The rational filter \( S_{gu}^{-1}(\theta) \) and the covariance matrix \( N(\theta) \) can be obtained from \( \theta \) using well established techniques, see Åström (1970) and Anderson and Moore (1979).

The above discussion indicates that \( \theta \) can be estimated using a prediction error method (PEM) applied to (1.5). The PEM estimate \( \hat{\theta} \) of \( \theta \) can be obtained as

\[ \hat{\theta} = \arg \min_\theta h(\theta) \]  \hspace{1cm} (1.6)

where \( h(\cdot) \) is a monotonically nondecreasing function defined on the domain of positive definite matrices, such as trace or determinant [see, e.g., Ljung (1976) for more details]. This approach to estimate \( \theta \) in the noisy input and output data case was thoroughly described in Söderström (1978, 1979, 1980, 1981), where it is shown that, while being quite involved computationally, it leads to accurate results (as expected, in view of the optimal PEM accuracy properties).

As is well known, when the number of data points \( N \) tends to infinity, the PE parameter estimate vector \( \hat{\theta} \) tends to the set

\[ \mathcal{P} = \{ \theta \in \Theta : S_{gu}(\theta) = \lambda(\theta) ; N(\theta) = N(\theta) \} \]  \hspace{1cm} (1.7)

see Ljung (1976). Since there is a one-to-one mapping between the class of input–output representations (1.5) having the above properties and the class of rational spectral density matrix
representations, it follows that \( \mathcal{S} \) can alternatively be defined as
\[
\mathcal{S} = \{ \Phi(z, \theta) \equiv \Phi(z, \theta) \}
\]  
(1.8)
where \( \Phi(z, \theta) \) is the spectral density matrix of \( \{ y(t), \sigma^2(t) \} \). The dependence of \( \Phi(z, \theta) \) on \( \theta \) is usually significantly less involved than that of \( \{ S(z, \theta), \Lambda(\theta) \} \). Therefore, the definition (1.8) of \( \mathcal{S} \) is more convenient than that of (1.7) when investigating the uniqueness properties of the PEM estimate \( \theta \).

Now, the estimate of \( \theta \) provided by any consistent method which uses second-order data will lie in \( \mathcal{S} \). Thus, while the authors concentrate on the PE method for the sake of illustration, it should be quite clear that the uniqueness (or parameter identifiability) conditions which they obtain will also apply to any other consistent estimation method based on second-order data.

Söderström (1978–1981) gave sufficient conditions on the system properties guaranteeing that
\[
\mathcal{S} = \{ \theta \}
\]  
(1.9)
Sufficient conditions for (1.9) to hold true were also obtained by Anderson and Deistler (1984) and Anderson (1985) under more general assumptions than those of Söderström (1978–1981) and of this paper. Here the authors derive necessary and sufficient conditions for (1.9) to hold. Furthermore, they characterize the systems for which \( \mathcal{S} \) contains more than one point, and show that usually they are not likely to be encountered in applications.

2. Main results
Consider the spectral matrix equation \( \Phi(z, \theta) = \Phi(z, \theta) \) corresponding to the system (1.1–1.3). This implies that
\[
S^2(z) = (s^2(z) + z^{-1}) S(z) S(z) + z^2 S(z) = \lambda^2 S(z) + z^2 S(z)
\]  
(2.1a)
\[
\frac{G(z)G(z^{-1})}{A(z)A(z^{-1})} \equiv \frac{\lambda^2}{z^2} + \frac{G(z)G(z^{-1})}{A(z)A(z^{-1})} \equiv \lambda^2 + \frac{G(z)G(z^{-1})}{A(z)A(z^{-1})}
\]  
(2.1b)
\[
\frac{G(z)G(z^{-1})}{A(z)A(z^{-1})} \equiv \lambda^2 + \frac{G(z)G(z^{-1})}{A(z)A(z^{-1})}
\]  
(2.1c)
Inserting (2.1c) into (2.1a) and (2.1b) gives, respectively,
\[
\lambda^2 S(z) + z^2 S(z) = \lambda^2 S(z) + z^2 S(z)
\]  
(2.2a)
\[
\lambda^2 S(z) + z^2 S(z) = \lambda^2 S(z) + z^2 S(z) \equiv \lambda^2 + \frac{G(z)G(z^{-1})}{A(z)A(z^{-1})}
\]  
(2.2b)
It follows from (2.2) that
\[
\lambda^2 S(z) + z^2 S(z) = \lambda^2 S(z) + z^2 S(z)
\]  
(2.3)
Omitting cases where \( B(z) \equiv 0 \), which clearly cannot lead to solutions of (2.1), (2.2), then if \( \lambda \neq \lambda^2 \) equation (2.3) implies (recall that \( u_{\alpha \neq 0}, b_{\alpha \neq 0} \) by assumption)
\[
\lambda^2 = \lambda^2; \quad \lambda^2 = \lambda^2.
\]  
(2.4)
Thus when \( \alpha \neq \lambda^2 \) it can easily be concluded from (2.4) and (2.1) that the only possible solution of (2.1) is \( \theta \).

To analyze cases for which other solutions than \( \theta \) are possible, in what follows it is assumed that \( \alpha = \lambda^2 \). Let \( A(z) = z^{-1} \) and similarly for \( B(z) \). Then (2.3) can be written as
\[
\lambda^2 = \lambda^2 = \lambda^2 = \lambda^2.
\]  
(2.5)
Clearly, (2.5) implies (2.4) if \( B(z) \equiv 0 \) for some \( z \) with \( |z| \geq 1 \), since then \( B(z) \) has some zeros inside or on the unit circle, and these zeros can neither be included in \( A(z) \) (since \( A(z) \neq 0 \) for \( |z| \leq 1 \) in the PEM) nor in the zeros of \( A^*(z) \) (since \( A^*(z) \) and \( B^*(z) \) are coprime by assumption).

In the following it is assumed that \( B(z) \neq 0 \) for \( |z| \geq 1 \). Then (2.5) implies either (2.4) or
\[
\lambda^2 = \lambda^2 = \lambda^2 = \lambda^2.
\]  
(2.6)
where \( \alpha = \lambda^2 \) is an arbitrary non-zero scalar. Next turn to (2.1).

From (2.1b) the following must hold:
\[
\mathcal{S}(z) = \mathcal{S}(z).
\]  
(2.7)
Inserting (2.6) and (2.7) into (2.1c), one obtains
\[
\lambda^2 G(z) G(z^{-1}) = \frac{\lambda^2 b_{\alpha \neq 0}}{\alpha} \frac{[B^*(z)/b_{\alpha \neq 0}] [B^*(z^{-1})/b_{\alpha \neq 0}]}{A(z)A(z^{-1})} \cdot G(z) G(z^{-1}).
\]  
(2.8)
Equation (2.8) has a solution if and only if
\[
\frac{b_{\alpha \neq 0}}{\alpha} > 0 \quad \text{and} \quad \frac{[B^*(z)/b_{\alpha \neq 0}] [B^*(z^{-1})/b_{\alpha \neq 0}]}{A(z)A(z^{-1})} \equiv P(z).
\]  
(2.9)
Where \( P(z) \) is a polynomial of order \( \eta \). Note that in view of the imposed condition that \( B(z) \neq 0 \) for \( |z| \geq 1 \), all the zeros of \( P(z) \) must lie outside the unit circle. Now assume that (2.9) holds. Then (2.8) implies
\[
\lambda^2 = \lambda^2 = \lambda^2 = \lambda^2.
\]  
(2.10)
Finally, inserting (2.7) and (2.10) into (2.1b) gives
\[
\lambda^2 = \lambda^2 = \lambda^2 = \lambda^2.
\]  
(2.11)
Thus it can be concluded that if the true system parameters satisfy (2.11) for some \( \alpha \) (with sign \( \alpha = \text{sign} b_{\alpha \neq 0} \) and some \( \lambda^2 \neq \lambda^2 \)), then there exists a false solution to (2.1), which is given by (2.6), (2.7) and (2.10).

Note that if \( \eta \gg \eta \) then (2.11) readily implies \( \lambda^2 = \lambda^2 \). \( \alpha = b_{\alpha \neq 0} \). Similarly, if the polynomials \( G(z) \) and \( P(z) \) have common zeros then (2.11) cannot hold for \( \lambda^2 = \lambda^2 \) (since \( G(z) \) and \( H(z) \) are coprime).

Based on the results above, the following theorem can now be stated.

**Theorem 2.1.** A necessary and sufficient condition for (2.1) to have a false solution (i.e. other than \( \theta \), the vector of true parameters) is that all the following conditions hold
(a) \( \eta = \eta \);
(b) \( B(z) \neq 0 \) for \( |z| \geq 1 \);
(c) \( G(z)B^*(z)/b_{\alpha \neq 0} \equiv A(z)P(z), P(z) \) being a polynomial;
(d) there exists a constant \( \alpha \neq 0 \) (with sign \( \alpha = \text{sign} b_{\alpha \neq 0} \) and \( \lambda^2 = \lambda^2 \) such that (2.11) is satisfied.

For (d) to hold it is necessary that \( \eta \gg \eta \), and \( P(z) \) and \( G(z) \) are coprime. Whenever the above conditions are satisfied, the corresponding false solution is given by (2.6), (2.7) and (2.10).

Some remarks are now in order.

(i) Sufficient conditions for (1.9) to hold can be obtained by violating any of the assumptions (a)–(d). For instance, any of the following conditions will guarantee (1.9)
\[
B(z) \neq 0 \quad \text{for} \quad |z| \leq 1;
\]  
(2.12a)
\[
A(z) \quad \text{and} \quad G(z) \quad \text{are coprime};
\]  
(2.12b)
\[
\eta > \eta.
\]  
(2.12c)
Such conditions have been derived by Söderström (1978, 1981).

(ii) Using Theorem 2.1 one can construct counterexamples to the uniqueness of PE parameter estimates for systems with noisy input and output data; see the next section. However, from the discussion above it should be quite clear that systems satisfying
Table 1. Parameter estimates of Example 3.1.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True values</th>
<th>Estimated values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Init. 1</td>
</tr>
<tr>
<td>$a_1$</td>
<td>0.5</td>
<td>0.519 ± (0.043)</td>
</tr>
<tr>
<td>$b_0$</td>
<td>1.0</td>
<td>0.956 ± (0.132)</td>
</tr>
<tr>
<td>$b_1$</td>
<td>4.0</td>
<td>4.073 ± (0.105)</td>
</tr>
<tr>
<td>$g_1$</td>
<td>0.5</td>
<td>0.547 ± (0.045)</td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>1.0</td>
<td>1.028 ± (0.075)</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>1.0</td>
<td>0.978 ± (0.033)</td>
</tr>
<tr>
<td>$\lambda_3$</td>
<td>1.0</td>
<td>0.981 ± (0.037)</td>
</tr>
<tr>
<td>Minimum PE loss</td>
<td>—</td>
<td>4.866 ± (0.068)</td>
</tr>
</tbody>
</table>

Conditions (a)--(d) are unlikely to occur, since their parameters must obey several non-trivial relationships.

3. A numerical example

Consider a system characterized by (1.1)–(1.3) and

$$
A(z) = 1 + 0.5z \\
B(z) = 1 + 4z \\
G(z) = 1 + 0.5z \\
H(z) = 1 \\
\lambda_1 = 1 \\
\lambda_2 = 1 \\
\lambda_3 = 1.
$$

(3.1)

Conditions (a)–(c) are clearly satisfied for this system [with $P(2) = 1 + 0.25z$]. Simple calculations show that condition (d) is also satisfied, with

$$
z = 2, \quad \lambda_3 = 0.125.
$$

It then follows from Theorem 2.1 that besides the true system parameter solution corresponding to (3.1), the spectral equations (2.1a)–(2.1c) also have a false solution given by

$$
\tilde{A}(z) = 1 + 0.25z \\
\tilde{B}(z) = 1 + 2z \\
\tilde{G}(z) = 1 + 0.25z \\
\tilde{H}(z) = 1 \\
\tilde{\lambda}_1 = 2 \\
\tilde{\lambda}_2 = 8 \\
\tilde{\lambda}_3 = 0.125.
$$

(3.2)

Thus, asymptotically for $N \to \infty$, the PE loss function in (1.6) associated with the special system (1.1)–(1.3), (3.1) has two global minima corresponding to (3.1) and (3.2), respectively. This should be true also for a "sufficiently large" (but finite) number $N$ of data points.

To verify the results above the authors have generated with the system (1.1)–(1.3), (3.1) a number of $N = 500$ data pairs. The PE algorithm outlined in Section 1 [also see Söderström (1981) for more details] was then applied to estimate the system parameters. Two different starting points were used in the algorithm. The first one, which will be called Init. 1, corresponds to the true solution in (3.1); the second one, Init. 2, corresponds to the false solution (3.2). The results for the means and standard deviations of the system parameter estimates as obtained from 10 independent runs are presented in Table 1. These results confirm the existence of a false global minimum point, as predicted by the asymptotic analysis developed in this paper.

4. Conclusion

The uniqueness of the PE estimate has been analyzed for a class of linear systems with noisy input and output data. Theorem 2.1 provides necessary and sufficient conditions for the corresponding PE loss function to have, asymptotically, more than one global minimum. A PE scheme may yield a false system parameter estimate whenever these conditions hold. The characteristics of the possible false solution have also been presented. However, the authors' analysis indicates that such systems are unlikely to occur in practice, since they have to obey several non-trivial relationships. Previously published sufficient conditions for a unique and true PE parameter estimate are obtained as a special case of the authors' analysis by violating any of the above conditions.

Finally, note two facts which albeit should be rather obvious. First, a system which satisfies the conditions of Theorem 2.1 is not uniquely identifiable by any method based on the second properties of the data. Thus the lack of uniqueness which occurs in such a case is not a peculiarity of the PEM but of the system itself. Second, it follows from Theorem 2.1 that a necessary and sufficient condition for uniqueness is that not all the relations (a)–(d) hold simultaneously. This statement in fact is an alternative formulation of Theorem 2.1.

Acknowledgements—The authors wish to acknowledge the contribution of Professor Torsten Söderström who provided the main part of the computer programs necessary for the numerical study reported in Section 3 and of Miss Carmen Martin who performed the numerical simulations.

References