An ARMA Robust System Identification Using a Generalized $l_p$ norm Estimation Algorithm

Bor-Sen Chen, Senior Member, IEEE, Jeng-Ming Chen, and Shen-Ching Shern

Abstract—The parameter estimation of ARMA systems with noisy output data is considered. The system output is corrupted by measurement noise, and the noise distribution is assumed to be unknown. A generalized $l_p$ norm iterative estimation algorithm ($1 < p < \infty$) is proposed to approach the maximum likelihood estimate of system parameters according to the noise distribution. Since the exponent $p$ of $l_p$ norm estimation algorithm is sensitive to the noise distribution, based on the sample kurtosis of the residual, an adequate exponent $p$ of $l_p$ norm estimation algorithm can be selected to achieve the efficient parameter estimation at each iteration step. Finally, several simulation results are presented to illustrate the proposed $l_p$ norm estimation algorithm; and we find that the proposed generalized $l_p$ norm estimation algorithm offers significant advantage to robust system parameter estimation problem with unknown noise distribution.

I. INTRODUCTION

The problem of system identification is a major field in control and signal processing. In general, the least squares is regarded as the most suitable method for estimating the coefficients in a regression model when the noises are Gaussian-distributed. Furthermore, the Gauss-Markoff theorem indicated that the least-squares estimator is the best linear unbiased estimator (BLUE) when the noise distribution is white-Gaussian [1]; but the least-squares regression is very far from optimal in many non-Gaussian situations [2]–[4]. Besides, Andrews [4] noted that even when the noise follows a Gaussian distribution, alternatives to least squares may be required, especially when the form of the system model is not exactly known. Therefore, if the quadratic loss function is not a suitable measure of the loss (for example, while the noise distribution is non-Gaussian), it is necessary to develop an alternative estimation approach.

However, an alternative to the least-squares estimator, namely, the class of $l_p$ norm estimators, has been proposed. Consider the linear equation

$$y(t) = X(t) \beta + e(t) \quad \forall t.$$  

(1)

Given a constant $p \geq 1$, the corresponding $l_p$ norm estimate $\hat{\beta}$ of the parameter vector $\beta$ is formulated as

$$\min_{\beta} \sum_{t=1}^{n} |y(t) - X(t)\hat{\beta}|^p \quad \text{for } 1 \leq p < \infty$$

$$\min_{\beta} \max_{t} |y(t) - X(t)\hat{\beta}| \quad \text{for } p = \infty.$$  

(2)

For some special cases, for example, the least squares (LS) ($p = 2$), the least absolute deviation (LAD) ($p = 1$), and the Chebychev (or minimax) estimator ($p = \infty$) methods have been proposed by many authors, e.g., [5]–[7]. It is well-known that the $l_1$ norm estimators are efficient when the noise follows the Laplace distribution, the $l_2$ norm estimators are efficient when the noise follows the Gaussian distribution, and the $l_\infty$ norm estimators are efficient when the noise follows the Uniform distribution. Furthermore, it has been shown that the $l_p$ norm estimator is the maximum likelihood estimator if noise is a generalized $p$-Gaussian (gpG) [8]; thus, we can always find a suitable exponent $p$ for the $l_p$ norm method, and its results are better than other linear methods for the systems with generalized $p$-Gaussian noise. Therefore, a priori information of noise distribution is necessary if we are to develop an efficient estimation.

Under the assumption that the true noise distribution is unknown, the purpose of this paper is to investigate an efficient parameter estimation method. As is well-known, the Cramér-Rao bound depends on the noise distribution, and the maximum likelihood estimation is asymptotically achievable to the Cramér-Rao bound [9]. Since estimate methods that are sufficient for an alternative noise distribution may be quite different, it is desirable that an estimation formula provide good estimates for a wide variety of unknown noise distributions. Therefore, a generalized $l_p$ norm method is proposed in this paper to solve the system parameter estimation problem with unknown output noise distribution. Furthermore, under the noisy output data case, if we apply the $l_p$ norm estimator directly, it will lead to a biased estimation. With the help of a whitening filter, an unbiased $l_p$ norm estimator can be obtained.

Because of a priori information of the noise distribution is assumed to be unknown and the exponent $p$ of the $l_p$ norm estimation algorithm is sensitive to the noise distribution, it is necessary to search for an adequate value of $p$ in each estimation step of the proposed $l_p$ norm iterative estimation algorithm. In fact, an adequate exponent $p$ can be selected via the calculation of the sample kurtosis of residual (i.e., the difference between the desired and the predicted), and these approaches have been proposed by several authors.
The paper is organized in the following way. In Section II, we formulate the problem of the ARMA system identification. In Section III, an $l_p$ norm estimation algorithm (iteratively reweighted least squares (IRLS)) is discussed. Furthermore, based on the Newton’s type method, we proposed an $l_p$ norm estimation algorithm—modified iteratively reweighted least squares (MIRLS) algorithm. In Section IV, we derive a generalized $l_p$ norm estimation algorithm based on the principle of whitening the correlated disturbance. The convergence analysis of the proposed $l_p$ norm estimation algorithm is shown in Section V. Finally, in Section VI, some simulation results are given to illustrate the proposed generalized $l_p$ norm estimation method.

II. PROBLEM FORMULATION

Consider a discrete-time linear system described by the following difference equation

$$A(q^{-1})z(t) = B(q^{-1})u(t)$$  \hspace{1cm} (3)

where

$$A(q^{-1}) = 1 + a_1q^{-1} + \cdots + a_nq^{-n}$$

$$B(q^{-1}) = b_0 + b_1q^{-1} + \cdots + b_nq^{-n}$$  \hspace{1cm} (4)

$a_{n_1}, b_{n_1} \neq 0$, and $q^{-1}$ is the backward shift operator. Without loss of generality, let $b_0 \neq 0$. At time $t$, $z(t)$ and $u(t)$ are, respectively, the output and input of the system (3). Suppose that all zeros of the polynomial $A(q^{-1})$ lie strictly inside the unit circle. Let $y(t)$ denote the noise corrupted measurement of $z(t)$, i.e.,

$$y(t) = z(t) + e(t)$$  \hspace{1cm} (5)

where the additive white noise $e(t)$ accounts for the random disturbance in the system as well as the measurement error. We assume that $e(t)$ is a zero mean stationary random process with symmetric distribution and autocorrelation function $\sigma^2(t)$. It is also assumed that $e(t)$ is uncorrelated with the input $u(t)$ or the output $z(t)$.

If the observation equation (5) is now substituted into (3), then we obtain the following noisy system model

$$A(q^{-1})y(t) = B(q^{-1})u(t) + n(t)$$  \hspace{1cm} (6)

where

$$n(t) = A(q^{-1})e(t).$$  \hspace{1cm} (7)

Note that $n(t)$ is a correlated random process. Thus (3) and (6) can be written in a matrix form as

$$Z = X_{z,u}\theta$$

and

$$Y = X_{y,u}\theta + N$$  \hspace{1cm} (8)

respectively, where we have (9), which appears at the bottom of the page. Given an equation like (8), the $l_p$ norm estimate of $\theta$ is obtained by minimizing

$$S' = ||Y - X_{y,u}\theta||_p = ||N||_p.$$  \hspace{1cm} (10)

For $p = 2$, the $l_2$ norm estimation (least squares) of the parameter $\theta$ is given by

$$\hat{\theta} = (X_{y,u}^TX_{y,u})^{-1}X_{y,u}^TY.$$  \hspace{1cm} (11)

Since $n(t)$ is non-white, we know that $\hat{\theta}$ is biased [1]. In fact, Harvey showed that when the disturbance term in the regression model is i.i.d. and with symmetric distribution, an $l_p$-norm estimator will always be unbiased for $1 < p < \infty$, provided that its first moment exists [14]. Now, from (7), we have

$$e(t) = \frac{1}{A(q^{-1})} \cdot n(t).$$  \hspace{1cm} (12)

Since $e(t)$ is a white noise with symmetric distribution, an unbiased $l_p$ norm estimation can be obtained by minimizing

$$S = ||e||_p$$

where

$$e = [e(t + 1) \cdots e(t + n)]^T.$$  \hspace{1cm} (9)

Thus, this result gives us an idea to develop an unbiased $l_p$ norm estimator.
III. MIRLS ESTIMATION ALGORITHM

Since the \( l_p \) norm method provides a good approach to the parameter estimation problem, in this section, we will discuss the \( l_p \) norm estimation method. First, we discuss the \( l_p \) norm parameter estimation algorithm under white-noise case, and the problem of \( l_p \) norm parameter estimation for system (8) with colored noise will be discussed in the next section. Consider the general linear system model as follows:

\[
y = X \beta + \epsilon
\]

where

- \( y \) \( n \)-vector of the observable random variables
- \( X \) \( n \)-by-\( k \) matrix of known regressor variables
- \( \beta \) \( k \)-vector of unknown parameters
- \( \epsilon \) \( n \)-vector of unobservable white noises.

The \( l_p \) norm estimation problem is then defined in the following:

Find the parameter vector \( \hat{\beta} = [\hat{b}_1, \hat{b}_2, \ldots, \hat{b}_k]^T \) that minimizes

\[
S(\hat{\beta}) = \sum_{i=1}^{n} |y_i - X_i \hat{\beta}|_p
\]

where \( X_i \) is the \( i \)-th row vector of matrix \( X \), \( y_i \) is the \( i \)-th element of \( y \). Let \( \epsilon = y - X \hat{\beta} \) be the corresponding \( n \)-vector of residuals (the estimate errors) and \( \epsilon_i \) be the \( i \)-th element of \( \epsilon \). Hence, \( \hat{\beta} \) is the \( l_p \) norm estimate of \( \beta = [b_1, b_2, \ldots, b_k]^T \).

The iteratively reweighted least squares (IRLS) algorithm is applicable to a variety of optimization problems in which some function of residuals \( \epsilon \) is minimized, e.g., [15]. For \( l_p \) norm minimization, the IRLS solves the problem represented in (14) by iteratively computing

\[
\hat{\beta}_{(k+1)} = (X^T W_{(k+1)}^{(k+1)} X)^{-1} X^T W_{(k+1)}^{(k+1)} y
\]

where the matrix \( W_{(k+1)}^{(k+1)} \) is a diagonal matrix whose \( i \)-th entry \( (W_{(k+1)}^{(k+1)})_{ii} = (W_{(k+1)}^{(k+1)})_{(k+1)} \) is calculated from the residual \( \epsilon_i^{(k)} = y_i - X_i \hat{\beta}_{(k)} \) (\( k \) \( \equiv \) iteration index), i.e.,

\[
W_{i}^{(k+1)} = [\epsilon_i^{(k)}]^p - 2.
\]

Thus, each iteration of the IRLS algorithm solves a new \( l_2 \) norm problem by employing the weighted residuals of the previous iteration in the current one.

Recognizing that, for \( 1 < p < 2 \), (16) produces a large value of \( W_{i}^{(k+1)} \) when the residual \( \epsilon_i^{(k)} \) is close to zero. Thus, in such cases, the IRLS algorithm does not meet its convergence condition: the weights \( W_{i}^{(k+1)} \) must be bounded for all \( \epsilon_i^{(k)} \) [16].

Therefore, for \( 1 < p < 2 \), \( W_{i}^{(k+1)} \) can be chosen as follows [16]

\[
W_{i}^{(k+1)} = \begin{cases} 
|\epsilon_i^{(k)}|^{p-2} & |\epsilon_i^{(k)}| \geq \epsilon \\
\epsilon^{p-2} & |\epsilon_i^{(k)}| < \epsilon
\end{cases}
\]

for a small positive number \( \epsilon \).

In general, the IRLS algorithm converges more slowly than the Newton’s type method. Therefore, in the following, we introduce an \( l_p \) norm estimation algorithm (MIRLS algorithm), which is a modified IRLS method by Newton–Raphson method to improve the convergent rate (see Fig. 1).

In fact, the minimum of \( S(\hat{\beta}) \) in (14) may be achieved by finding the roots of

\[
\frac{\partial S(\hat{\beta})}{\partial \hat{b}_i} = 0 \quad i = 1, 2, \ldots, k
\]

whereas (18) can be solved iteratively by finding the solution of

\[
\frac{\partial^2 S(\hat{\beta})}{\partial \hat{b}_i \partial \hat{b}_j} - \sum_{j=1}^{n} \frac{\partial S(\hat{\beta})}{\partial \hat{b}_j} \frac{\partial S(\hat{\beta})}{\partial \hat{b}_j} = 0
\]

which is the Newton–Raphson method [17].

Hence, from (14), we find that (at the \( n \)-th iteration)

\[
\frac{\partial S(\hat{\beta})}{\partial \hat{b}_i} = \frac{\partial^2 S(\hat{\beta})}{\partial \hat{b}_i \partial \hat{b}_j} = \frac{\partial^2 S(\hat{\beta})}{\partial \hat{b}_j \partial \hat{b}_j} = 0
\]

where \( \epsilon_i = \text{sign}(y_i - X_i \hat{\beta}(k-1)) \).

Therefore, the algorithm converges when

\[
\sum_{i=1}^{n} |y_i - X_i \hat{\beta}(k)|_p = 0
\]

Finally, we can conclude that the MIRLS algorithm converges when

\[
\sum_{i=1}^{n} |y_i - X_i \hat{\beta}(k)|_p = 0
\]

Fig. 1. Flowchart of the generalized \( l_p \) norm estimation algorithm.
\( W_i^{(\kappa)} = [y_i - X_i \hat{\beta}^{(\kappa-1)}] p^{-2} \), and \( X_i \) is the \( i\)th entry of matrix \( X \).

Substituting (20) into (19), we have

\[
(p - 1)H^{(\kappa)}(\hat{\beta}^{(\kappa)} - \hat{\beta}^{(\kappa-1)}) = D^{(\kappa)} - H^{(\kappa)} \hat{\beta}^{(\kappa-1)}
\]

where \( H^{(\kappa)} \) is a k-by-k matrix with \( \sum_{i=1}^n W_i^{(\kappa)} X_i X_i \) in \( i\)th entry and

\[
D^{(\kappa)} = \left[ \sum_{i=1}^n W_i^{(\kappa)} y_i X_i, \ldots, \sum_{i=1}^n W_i^{(\kappa)} y_i X_n \right]^T.
\]

Note that \( H^{(\kappa)} = X^T W^{(\kappa)} X \) and \( D^{(\kappa)} = X^T W^{(\kappa)} y \); then, by (15), we get \( H^{(\kappa)} \hat{\beta}^{(\kappa)}_{\text{IRLS}} = D^{(\kappa)} \). Therefore, we have

\[
(p - 1)H^{(\kappa)}(\hat{\beta}^{(\kappa)} - \hat{\beta}^{(\kappa-1)}) = H^{(\kappa)}(\hat{\beta}^{(\kappa)}_{\text{IRLS}} - \hat{\beta}^{(\kappa-1)}).
\]

It is noted that the matrix \( H^{(\kappa)} \) is nonsingular. Thus, we get the following algorithm

\[
\hat{\beta}^{(\kappa)} = \frac{((p - 2)\hat{\beta}^{(\kappa-1)} + \hat{\beta}^{(\kappa)}_{\text{IRLS}})}{p - 1}.
\] (21)

We know that the above algorithm is a Newton's type method, and it is a rearranged form of Newton-Raphson method. Since the Newton's method is theoretically the most desirable, in this paper, the iteration scheme of (21) plays a major role in the proposed \( l_p \) norm estimation algorithm for finding the roots of (18).

Remarks: (i) In general, the algorithm (21) based on the Newton's type method converges faster than IRLS algorithm. However, we know that Newton's type method does not always converge for \( 1 < p \leq \frac{3}{2} \). Thus, \( p = \frac{3}{2} \) is indeed the critical value for Newton's type method [18]. Therefore, for avoiding the divergence to occur, the algorithm in (21) is used only for \( \frac{3}{2} < p < \infty \) case. For \( 1 < p \leq \frac{3}{2} \) case, the IRLS algorithm in (15) is still used. For further discussion, we will investigate the convergence properties of the estimation algorithm (21) in the next section. (ii) If the sequence \( \{\beta^{(\kappa)}\} \) in algorithm (21) converges to \( \beta^* \) and \( \{\hat{\beta}^{(\kappa)}_{\text{IRLS}}\} \) in the IRLS algorithm converges to \( \beta_{\text{IRLS}} \), by (21), we have

\[
E[\beta^*] = \frac{((p - 2)E[\hat{\beta}^*] + E[\hat{\beta}_{\text{IRLS}}])}{p - 1}
\]

i.e.,

\[
E[\beta^*] = E[\hat{\beta}_{\text{IRLS}}].
\] (22)

Then, by (22), we know that if the IRLS algorithm is unbiased (or biased), the algorithm (21) based on Newton-Raphson method is also unbiased (or biased).

Now based on the algorithm in (21) for \( \frac{3}{2} < p < \infty \) case and the IRLS algorithm in (15) for \( 1 < p \leq \frac{3}{2} \) case, we propose the following \( l_p \) (\( 1 < p < \infty \)) norm estimation algorithm in which \( j \) denotes the iteration step.

**Algorithm (MIRLS)**

**Step 0:** Give a small positive number \( \epsilon \) and an initial estimate \( \hat{\beta}^{(j)} \) of \( \beta \) at \( j = 1 \).

**Step 1:** Calculate the residuals

\[
\epsilon_i^{(j)} = y_i - X_i \hat{\beta}^{(j)}
\]

**Step 2:** Set

\[
W_i^{(j + 1)} = \begin{cases} 
\epsilon_i^{(j)} & \text{if } 1 < p < \infty \\
\min(\epsilon_i^{(j)} p^{-2}, \epsilon p^{-2}) & \text{if } 1 < p < 2
\end{cases}
\]

**Step 3:** Find the IRLS estimate of \( \beta \)

\[
\hat{\beta}^{(j + 1)}_{\text{IRLS}} = (X^T W^{(j + 1)} X)^{-1} X^T W^{(j + 1)} y
\]

**Step 4:** Modify the IRLS estimation \( \hat{\beta}^{(j + 1)}_{\text{IRLS}} \) by the following formula

\[
\hat{\beta}^{(j + 1)} = \begin{cases} 
\hat{\beta}^{(j + 1)}_{\text{IRLS}} & \text{if } 1 < p \leq \frac{3}{2} \\
\frac{((p - 2)\hat{\beta}^{(j)} + \hat{\beta}^{(j + 1)}_{\text{IRLS}})}{p - 1} & \text{if } \frac{3}{2} < p < \infty
\end{cases}
\]

(23)

**Step 5:** Let \( j = j + 1 \). If \( \hat{\beta} \) converges then STOP, else go the Step 1.

IV. GENERALIZED \( l_p \) NORM ESTIMATION ALGORITHM

A. Generalized \( l_p \) Norm Estimation Algorithm

Now that \( l_p \) norm estimation algorithm MIRLS was introduced in the section above, a generalized \( l_p \) norm estimation algorithm for the system (8) with colored noise will be proposed in this section.

Fundamentally, the key for derivation of the generalized \( l_p \) norm estimation algorithm is that we introduce a whitening filter to convert the correlated noise \( n(t) \) in (6) into a white noise. To implement this idea, we assume that \( n(t) \) has a rational power spectrum and the following autoregressive model is satisfied.

\[
n(t) + \sum_{i=1}^{p_c} c_i n(t - i) = e(t)
\]

(24)

where \( c_i \) for \( i = 1, \ldots, p_c \) are constant coefficients and \( p_c \) is the order of the model.

Rewrite (24) as

\[
C(q^{-1}) n(t) = e(t)
\]

(25)

where

\[
C(q^{-1}) = 1 + c_1 q^{-1} + \cdots + c_{p_c} q^{-p_c}
\]

(26)

Thus we can get the following system model by combining (6) and (25):

\[
C(q^{-1}) A(q^{-1}) y(t) = C(q^{-1}) B(q^{-1}) u(t) + e(t)
\]

(27)

This model now has a white noise \( e(t) \). Then, unbiased parameter estimates of the coefficients of the polynomials
System (30) is similar to system (13), and $e(t)$ is a white noise, by [14], an unbiased $l_p$ norm parameter estimate of $\theta_c$ can be obtained by minimizing the following error function

$$S_c = \sum_t |n(t) - \hat{\varphi}(t)\hat{\theta}_c|^p$$

where $\hat{\theta}_c = [\hat{\varphi}_1 \cdots \hat{\varphi}_p]^T$ is the estimate of $\theta_c$.

Therefore, the proposed MIRLS algorithm in the section above is used to estimate the parameters $c_i$ by minimizing $S_c$.

Now, based on the MIRLS algorithm, we can develop the following $l_p(1 < p < \infty)$ norm estimation algorithm to estimate the parameters $a_i$, $b_i$, and $c_i$ by minimizing the error function $S_p$ and $S_c$, respectively (in which $j$ denotes the iteration step), and a flow chart of the generalized $l_p$ norm estimation algorithm is given in Fig. 1.

Main Algorithm

**Step 0:** (initial estimate) Let $j = 0$, $\hat{C}(j)(q^{-1}) = 1$, and $p(j) = 2$. Use the $l_2$ norm method and bias reduction technique [20] to find $\hat{A}(j)(q^{-1})$ and $\hat{B}(j)(q^{-1})$.

**Step 1:** With estimates $\hat{A}(j)(q^{-1})$ and $\hat{B}(j)(q^{-1})$, calculate the residual $n(t)$ by

$$\hat{u}(t) = \hat{A}(j)(q^{-1})\hat{y}(t) - \hat{B}(j)(q^{-1})n(t).$$

Select an exponent $p(j+1)$ from $\hat{C}(j)(q^{-1})\hat{u}(t)$, then use the corresponding $l_p$ norm method (MIRLS) to find $\hat{C}(j+1)(q^{-1})$ by minimizing

$$S_c = \sum_t |\hat{C}(j+1)(q^{-1})\hat{u}(t)|^p.$$

(Remark: The method of selecting the exponent $p(j+1)$ will be introduced in subsection B.)

**Step 2:** With estimate $\hat{C}(j+1)(q^{-1})$, calculate two filtered signals $\hat{u}(t)$ and $\hat{y}(t)$ as

$$\hat{u}(t) = \hat{C}(j+1)(q^{-1})\hat{u}(t), \hat{y}(t) = \hat{C}(j+1)(q^{-1})\hat{y}(t).$$

Select an exponent $p(j+1)$ from $\hat{C}(j+1)(q^{-1})\hat{u}(t)$, then use the corresponding $l_p$ norm method (MIRLS) to find $\hat{A}(j+1)(q^{-1})$ and $\hat{B}(j+1)(q^{-1})$ by minimizing

$$S_p = \sum_t |\hat{A}(j+1)(q^{-1})\hat{y}(t) - \hat{B}(j+1)(q^{-1})\hat{u}(t)|^p.$$

**Step 3:** Let $j = j + 1$. Go to Step 1 until $\hat{A}(q^{-1})$, $\hat{B}(q^{-1})$, and $\hat{C}(q^{-1})$ converge.

Remarks: (i) Since the generalized $l_p$ norm algorithm is an iterative scheme for solving the highly nonlinear minimization problem defined by (28), the convergence of the algorithm to the optimal solution is not always guaranteed. The typical problem of this case is that more than one minimum point in $S$ may exist. Thus, the choice of an initial parameter estimate near the optimal solution is crucial to find the global minimum successfully [7]. In this paper, the $l_2$ norm estimation is used for initial parameter estimation. However, from (7) we know that $n(t)$ is a correlated random variable; thus, in the initial estimate, the $l_2$ solution is biased. Therefore, we need an effective bias-reduction method to get a better initial
estimate. Such a method was proposed by the Le and Wilson [20]. They dealt with bias-reduction problem via the condition number based on the singular value decomposition (SVD).

(ii) Since \( e(t) \) is a white noise with symmetric distribution, we know that the MIRLS estimation algorithm of estimating the parameters \( a_i, b_i \) (or \( c_i \)) by minimizing the error functions \( S_p \) (or \( S_0 \)) is an unbiased algorithm [14]. Furthermore, if \( e(t) \) is a generalized \( p \)-Gaussian white noise, the MIRLS estimation is also the maximum likelihood estimation [8]. We can extend the discussion on the generalized \( l_p \) norm estimation algorithm to an unbiased estimation algorithm because the generalized \( l_p \) norm estimation is based on the MIRLS algorithm. In the next section, we will discuss the convergence properties of the generalized \( l_p \) norm estimation algorithm. Moreover, by (12) and (25), we know that the transfer function of the whitening filter is \( 1/A(q^{-1}) \), and the ideal order of \( C(q^{-1}) = (1/A(q^{-1})) \) is infinite. However, since the system model is assumed to be an asymptotically stable filter, the whitening filter transfer function \( C(q^{-1}) \) may be approximated arbitrarily by a polynomial with proper order \( p_c \). (iii) In fact, for \( l = 2 \) case, the approach to parameter estimation of minimizing the error function \( S \) in (28) is called the generalized least-squares (GLS) method. This method is an extension of the well-known least-squares (\( l_2 \)) method. Furthermore, the GLS estimates are equivalent to the Markov estimates [7]. In general, the \( l_2 \) method is regarded as the most suitable method for estimating the coefficients in a regression model when the noises are Gaussian-distributed. However, in many practical situations, the noises are not Gaussian distributed, and in such cases, the \( l_2 \) technique may provide relatively poor estimates of the regression coefficients. Basically, from the simulation results of Money et al. [10], we know that the \( l_p \) norm method provides alternatives to the least squares (\( l_2 \)) method for estimating the coefficients of a linear regression model. Furthermore, the \( l_p \) norm method \((p \neq 2)\) showed the improvement of efficiency over the least-squares method for all non-Gaussian noise distributions. Therefore, for system model (6) with unknown noise distribution, the generalized \( l_p \) norm estimation algorithm is a good approach to estimate the regression coefficients.

B. The Choice of the Exponent \( p \)

In the choice of exponent \( p \) in \( l_p \) norm estimation, criteria based on the kurtosis of the noise distribution have been proposed, e.g., [10]-[13]. Among them, Money et al. [10] indicate that the following empirical formula:

\[
p = \frac{9}{\beta_2^2} + 1
\]

(31)

(where \( \beta_2 \) is the kurtosis of the noise distribution) can be used in practice to determine a suitable \( p \), provided the noise distribution is symmetric. In particular, is should be noted that, for the Gaussian distribution (kurtosis \( \beta_2 = 3 \)), this formula suggests the use of \( p = 2 \).

The kurtosis \( \beta_2 \) of some distributions are given in Table I.

However, it is stressed that the investigation of Money et al. is based on the known noise distribution. In other words, \textit{a priori} information of kurtosis \( \beta_2 \) is necessary before applying (31) to the choice of \( p \). In many practical situations, the kurtosis of the noise distribution is unknown, and it must be estimated from the sample data. Therefore, in such cases, Money et al. [10] suggest that the true kurtosis can be replaced in (31) by a sample estimate of the true kurtosis to obtain a value of \( p \), which can be used to determine the \( l_p \) estimates of the regression coefficients.

Since the true distribution of measurement noise \( e(t) \) in (5) is assumed to be unknown, the kurtosis of noise distribution must be estimated, i.e., we need to estimate the value of exponent \( p \) before we use the MIRLS estimation method. We indicate the following method to approach the value of the exponent \( p \) from noisy observations with the help of Money’s formula (31).

1) Estimate the sample kurtosis of residual \( e(t) \)

\[
\beta_2^2 = \frac{N}{\sum_{t} (e(t) - \bar{e})^4} \left\{ \sum_{t} (e(t) - \bar{e})^2 \right\}^2
\]

where \( \bar{e} \) is the mean of \( e(t) \) and \( N \) is the sample size.

2) Then we select the \( \tilde{p} \) as

\[
\tilde{p} = \frac{9}{\beta_2^2} + 1.
\]

(32)

Remark: In general, for the sample size \( N = 400 \) case, the sample kurtosis \( \beta_2^2 \) provides a good approximation of the true kurtosis [12]. Therefore, in this paper, we use the sample size \( N = 400 \) in the simulation studies.

Fundamentally, the proposed generalized \( l_p \) norm estimation algorithm consists of an inner iteration (calculation of \( \tilde{p} \)) and an outer iteration (minimization over the parameters). Therefore, the limit of the sequence \( \{ \tilde{p}^{(i)} \} \) in the main algorithm can be obtained by the iteration calculation. Does the sequence \( \{ \tilde{p}^{(i)} \} \) converge? In the next section, we will discuss this problem. Moreover, in the appendix, we discuss the convergence property of \( \tilde{p} \) for the Gaussian distribution case.

V. CONVERGENCE ANALYSIS

In this section, we will analyze the convergence property of the proposed generalized \( l_p \) norm estimation algorithm. We first state a theorem.

Theorem: [21] Consider the following sequential equation

\[
X^{(i+1)} = BX^{(i)} + b \quad i = 0, 1, 2, \cdots
\]

(33)

where \( B \) is an \( n \)-by-\( n \) matrix and \( X^{(i)}, b \) are \( n \)-by-1 vectors.
Let $X^*$ be a point of attraction of the iteration (33). For any initial vector $X^{(0)}$, the vector sequence $\{X^{(n)}\}$ converges to a unique fixed point $X^*$ if and only if the following condition is satisfied

$$\max_{1 \leq j \leq n} |\lambda_j(B)| < 1$$  \hspace{1cm} (34)

where $\lambda_j(B)$ denotes an eigenvalue of $B$.

We now proceed to discuss the convergence properties of the proposed $l_p$ norm estimation algorithm. The following steps are quite similar to those in [22].

Taking the mathematical expectation of both sides of (21), we get

$$E[\hat{\beta}^{(j)}] = \frac{p-2}{p-1} E[\hat{\beta}^{(j-1)}] + \frac{1}{p-1} E[\hat{\beta}^{\text{IRLS}}]$$

Since the IRLS estimate $\hat{\beta}^{(j)}_{\text{IRLS}}$ in each iteration step is obtained by minimizing the error function $S_p$ (or $S_\tau$), by (14), we know that $E[\hat{\beta}^{(j)}_{\text{IRLS}}] = \theta$ for any iteration step $j$.

Then we get

$$E[\hat{\beta}^{(j)}] = \frac{p-2}{p-1} E[\hat{\beta}^{(j-1)}] + \frac{1}{p-1} \theta$$  \hspace{1cm} (35)

i.e.

$$E[\hat{\beta}^{(j)} - \theta] = \frac{p-2}{p-1} E[\hat{\beta}^{(j-1)} - \theta]$$

From the analogy between (33) and (35), we therefore deduce that the sequence $\{\hat{\beta}^{(j)}\}$ is convergent in the mean if the following condition is satisfied

$$\left|\frac{p-2}{p-1}\right| < 1$$

i.e.

$$\frac{3}{2} < p$$

Thus, for $\frac{3}{2} < p < \infty$, we can expect that the sequence $\{\hat{\beta}^{(j)}\}$ in (23) is convergent. In fact, it is well-known that, in (33), the matrix operator $B$ with $|\lambda(B)| < 1$ is referred to as a contraction mapping [21]. Hence, for $\frac{3}{2} < p < \infty$, the sequence $\{\hat{\beta}^{(j)}\}$ in (23) may be expected to be convergent and can never diverge. Moreover, the IRLS estimation algorithm is convergent if the weighting function $W_t^{(j)}$ is chosen as (16) or (17). Therefore, for $1 < p < \infty$, the proposed MIRLS estimation algorithm is convergent. Since the generalized $l_p$ norm estimation algorithm is based on the MIRLS algorithm, we can also expect that the generalized $l_p$ norm estimation algorithm is convergent. Now, we consider the unbiased property of the proposed generalized $l_p$ norm estimation algorithm. Harcey [14] showed that if the disturbance term in the regression model is symmetrically distributed, an $l_p$ norm estimator will always be unbiased for $1 < p < \infty$, provided that its first moment exists.

Since the measurement noise $e(t)$ is a white noise with symmetric distribution, an $l_p$ norm parameter estimate of $\theta$ (or $\theta_r$) by minimizing the error function $S_p$ (or $S_\tau$) will be unbiased. Then the IRLS algorithm of estimating the parameters $a_i, b_i$ (or $c_i$) is an unbiased estimation algorithm and, by (22), the MIRLS algorithm is also an unbiased estimation algorithm. Furthermore, Monev et al. [10] showed that, so far, studies using symmetric noise distributions have produced no evidence of bias in the $l_p$ norm estimates obtained using different values of $p$. In fact, based on the known noise distribution, the experimental results of Monev et al. showed that the $l_p$ norm estimation algorithm is always unbiased, and the efficiency of $l_p$ norm estimation algorithm will be improved by the choice of suitable $p$ value; therefore, it is concluded that the $l_p$ norm estimators are unbiased for all $1 < p < \infty$ when the noise distribution is symmetric. From the above, we conclude that even if the value of $p$ is updated in each iteration step, the MIRLS algorithm is always an unbiased estimation algorithm, and the proposed generalized $l_p$ norm estimation algorithm can be expected to be unbiased.

Note that the choice of $p$ value is based on an estimate of the kurtosis. Since the proposed MIRLS estimation algorithm is convergent, it is reasonable that we may expect the sequence $\{\hat{\beta}^{(j)}\}$ to be convergent. Furthermore, the proposed generalized $l_p$ norm estimation algorithm is an unbiased estimator, and $\hat{\beta}$ will converge to a suitable value. As will be examined in the next section, the simulation studies have shown that the proposed generalized $l_p$ norm estimation algorithm is an unbiased estimator, and $\hat{\beta}$ will converge to 2 for the Gaussian case.

Remark: In this paper, the generalized $l_p$ norm estimation for $1 < p < \infty$ case is considered. Moreover, the least-squares ($l_2$) method is efficient when the noise follows the Gaussian distribution. Thus, in the appendix, we will study the convergence property of $\hat{\beta}$ for the Gaussian case.

VI. SIMULATION RESULTS

In this section, we use the generalized $l_p$ norm estimation algorithm indicated in the previous section to solve the ARMA system identification problems. Two examples of ARMA system are illustrated with three different types of noise. In our simulations, the input driving sequence $u(t)$ is white-Gaussian with zero mean and unit variance. The data length is 400, and the order of whitening filter $p_r$ is 10.

Example 1: (ARMA(2,2), minimum-phase)

The system model is given by

$$z(t) - 1.5z(t-1) + 0.8z(t-2) = u(t) - 1.2u(t-1) + 0.9u(t-2)$$

$$y(t) = z(t) + e(t).$$

For the convenience of simulation, three types of disturbance $e(t)$ are given as follows:

1) Zero mean white Gaussian noise.

2) White Gaussian noise with zero mean and variance 0.1 multiplied by a white noise uniformly distributed in the interval $[-0.5, 0.5]$.

3) Zero mean white contaminated Gaussian noise with distribution of the form $f = (1/2)N(0, \sigma^2_e) + (1/2)N(0, 8\sigma^2_e)$. 

TABLE II

<table>
<thead>
<tr>
<th>Method</th>
<th>Parameter</th>
<th>$a_1$</th>
<th>$a_2$</th>
<th>$b_0$</th>
<th>$b_1$</th>
<th>$b_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>$-1.5$</td>
<td>$0.8$</td>
<td>$1.0$</td>
<td>$-1.2$</td>
<td>$0.9$</td>
<td></td>
</tr>
<tr>
<td>1 Mean</td>
<td>$-1.4821$</td>
<td>$0.7876$</td>
<td>$1.0044$</td>
<td>$-1.1857$</td>
<td>$0.8939$</td>
<td></td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>$0.0480$</td>
<td>$0.0381$</td>
<td>$0.0238$</td>
<td>$0.0414$</td>
<td>$0.0255$</td>
<td></td>
</tr>
<tr>
<td>2 Mean</td>
<td>$-1.4819$</td>
<td>$0.7860$</td>
<td>$1.0020$</td>
<td>$-1.1816$</td>
<td>$0.8912$</td>
<td></td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>$0.0479$</td>
<td>$0.0381$</td>
<td>$0.0184$</td>
<td>$0.0414$</td>
<td>$0.0227$</td>
<td></td>
</tr>
<tr>
<td>3 Mean</td>
<td>$-1.4854$</td>
<td>$0.7819$</td>
<td>$1.0047$</td>
<td>$-1.1888$</td>
<td>$0.8882$</td>
<td></td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>$0.0581$</td>
<td>$0.0442$</td>
<td>$0.0182$</td>
<td>$0.0550$</td>
<td>$0.0273$</td>
<td></td>
</tr>
<tr>
<td>4 Mean</td>
<td>$-1.5315$</td>
<td>$0.8556$</td>
<td>$1.0004$</td>
<td>$-1.2356$</td>
<td>$0.9515$</td>
<td></td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>$0.2110$</td>
<td>$0.2684$</td>
<td>$0.0119$</td>
<td>$0.2126$</td>
<td>$0.2096$</td>
<td></td>
</tr>
</tbody>
</table>

Method 1: The generalized $l_p$ norm estimation algorithm.
Method 2: The generalized $l_p$ ($p = 2$) norm estimation algorithm (GLS).
Method 3: The generalized $l_p$ ($p = 2.5$) norm estimation algorithm.
Method 4: The Bias-correction estimation algorithm.

TABLE III

<table>
<thead>
<tr>
<th>Method</th>
<th>Parameter</th>
<th>$a_1$</th>
<th>$a_2$</th>
<th>$b_0$</th>
<th>$b_1$</th>
<th>$b_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>$-1.5$</td>
<td>$0.8$</td>
<td>$1.0$</td>
<td>$-1.2$</td>
<td>$0.9$</td>
<td></td>
</tr>
<tr>
<td>1 Mean</td>
<td>$-1.4851$</td>
<td>$0.7884$</td>
<td>$0.9998$</td>
<td>$-1.1873$</td>
<td>$0.8991$</td>
<td></td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>$0.0226$</td>
<td>$0.0170$</td>
<td>$0.0136$</td>
<td>$0.0220$</td>
<td>$0.0128$</td>
<td></td>
</tr>
<tr>
<td>2 Mean</td>
<td>$-1.4819$</td>
<td>$0.7844$</td>
<td>$1.0044$</td>
<td>$-1.1812$</td>
<td>$0.8979$</td>
<td></td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>$0.0533$</td>
<td>$0.0432$</td>
<td>$0.0208$</td>
<td>$0.0501$</td>
<td>$0.0312$</td>
<td></td>
</tr>
<tr>
<td>3 Mean</td>
<td>$-1.4859$</td>
<td>$0.7825$</td>
<td>$1.0063$</td>
<td>$-1.1824$</td>
<td>$0.8893$</td>
<td></td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>$0.0522$</td>
<td>$0.0391$</td>
<td>$0.0204$</td>
<td>$0.0475$</td>
<td>$0.0253$</td>
<td></td>
</tr>
<tr>
<td>4 Mean</td>
<td>$-1.5495$</td>
<td>$0.8533$</td>
<td>$1.0018$</td>
<td>$-1.2514$</td>
<td>$0.9458$</td>
<td></td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>$0.1957$</td>
<td>$0.2738$</td>
<td>$0.0102$</td>
<td>$0.1933$</td>
<td>$0.2164$</td>
<td></td>
</tr>
</tbody>
</table>

Method 1: The generalized $l_p$ norm estimation algorithm.
Method 2: The generalized $l_p$ ($p = 2$) norm estimation algorithm (GLS).
Method 3: The generalized $l_p$ ($p = 2.5$) norm estimation algorithm.
Method 4: The Bias-correction estimation algorithm.

The SNR is 20 dB in this example. Averaging all the estimated results over 50 simulations, the results are given in Tables II–IV, which show the averaged solution and one standard derivation (Std. Dev.). The accuracy is chosen to be within ±10−6. To illustrate the performance of the proposed generalized $l_p$ norm estimation algorithm, Tables II–IV compare the statistics of parameter estimates obtained by using this approach with some methods including GLS and Bias-correction methods [7]. For the generalized $l_p$ norm estimation algorithm, the averaged estimate values of $p$ are, respectively, 1.97, 1.40, and 1.45 in type (i), (ii), and (iii) disturbances; and the average number of iteration steps are, respectively, 11, 26, and 21 in type (i), (ii), and (iii) disturbances. In general, the generalized $l_p$ norm estimation algorithm converges somewhat faster than the Bias-correction algorithm, but more slowly than GLS and $l_p$ ($p = 2.5$) algorithm.

Moreover, from the simulation results, it is recognized that the value of $p$ converges to 2 in the Gaussian case; and the $l_p$ norm estimators are unbiased for all $p$ and for all noise distributions. Thus, the results reveal that the generalized $l_p$ norm estimation algorithm is more efficient than any other unbiased estimator with the exception of the Gaussian case. It is well-known that the GLS estimation method is efficient when the noise distribution is Gaussian. However, the generalized $l_p$ norm estimation algorithm is still a good approach to estimate the system parameters in the Gaussian case. Therefore, it is believed that the proposed $l_p$ norm estimation algorithm is a good approach to the parameter estimation problem with unknown measurement noise distribution.

Example 2: (ARMA(2,1), nonminimum-phase)
The system model is given by

$$z(t) = -1.5z(t-1) + 0.8z(t-2) = u(t) - 1.25u(t-1)$$

$$y(t) = z(t) + e(t).$$

Similarly, for the convenience of simulation, three types of disturbance $e(t)$ are given as follows:

1) Zero mean white Gaussian noise.
2) White Gaussian noise with zero mean and variance 0.01 multiplied by a white noise uniformly distributed in the interval $[-0.5, 0.5]$.
3) Zero mean white contaminated Gaussian noise with distribution of the form $f = (1/2)N(0, \sigma^2_e) + (1/2)N(0, 8\sigma^2_e)$. 


TABLE IV

<table>
<thead>
<tr>
<th>Method</th>
<th>Parameter</th>
<th>$a_1$</th>
<th>$a_2$</th>
<th>$b_0$</th>
<th>$b_1$</th>
<th>$b_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>$-1.5$</td>
<td>$0.8$</td>
<td>$1.0$</td>
<td>$-1.2$</td>
<td>$0.9$</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Mean</td>
<td>$-1.4895$</td>
<td>$0.7897$</td>
<td>$1.0004$</td>
<td>$-1.1910$</td>
<td>$0.9005$</td>
</tr>
<tr>
<td></td>
<td>Std. Dev.</td>
<td>$0.0089$</td>
<td>$0.0049$</td>
<td>$0.0110$</td>
<td>$0.0144$</td>
<td>$0.0089$</td>
</tr>
<tr>
<td>2</td>
<td>Mean</td>
<td>$-1.4832$</td>
<td>$0.7880$</td>
<td>$1.0060$</td>
<td>$-1.1897$</td>
<td>$0.8944$</td>
</tr>
<tr>
<td></td>
<td>Std. Dev.</td>
<td>$0.0429$</td>
<td>$0.0351$</td>
<td>$0.0186$</td>
<td>$0.0381$</td>
<td>$0.0217$</td>
</tr>
<tr>
<td>3</td>
<td>Mean</td>
<td>$-1.4832$</td>
<td>$0.7883$</td>
<td>$1.0096$</td>
<td>$-1.1813$</td>
<td>$0.8931$</td>
</tr>
<tr>
<td></td>
<td>Std. Dev.</td>
<td>$0.0430$</td>
<td>$0.0341$</td>
<td>$0.0207$</td>
<td>$0.0394$</td>
<td>$0.0228$</td>
</tr>
<tr>
<td>4</td>
<td>Mean</td>
<td>$-1.5580$</td>
<td>$0.8524$</td>
<td>$0.9978$</td>
<td>$-1.2544$</td>
<td>$0.9529$</td>
</tr>
<tr>
<td></td>
<td>Std. Dev.</td>
<td>$0.2241$</td>
<td>$0.3217$</td>
<td>$0.0096$</td>
<td>$0.2212$</td>
<td>$0.2470$</td>
</tr>
</tbody>
</table>

Method 1: The generalized $l_p$ norm estimation algorithm.
Method 2: The generalized $l_p$ norm estimation algorithm (GLS).
Method 3: The generalized $l_p$ ($p = 2.5$) norm estimation algorithm.
Method 4: The Bias-correction estimation algorithm.

TABLE V

<table>
<thead>
<tr>
<th>Method</th>
<th>Parameter</th>
<th>$a_1$</th>
<th>$a_2$</th>
<th>$b_0$</th>
<th>$b_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>$-1.5$</td>
<td>$0.8$</td>
<td>$1.0$</td>
<td>$-1.25$</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Mean</td>
<td>$-1.4852$</td>
<td>$0.7872$</td>
<td>$1.0147$</td>
<td>$-1.2552$</td>
</tr>
<tr>
<td></td>
<td>Std. Dev.</td>
<td>$0.0122$</td>
<td>$0.0102$</td>
<td>$0.0250$</td>
<td>$0.0288$</td>
</tr>
<tr>
<td>2</td>
<td>Mean</td>
<td>$-1.4857$</td>
<td>$0.7891$</td>
<td>$1.0147$</td>
<td>$-1.2558$</td>
</tr>
<tr>
<td></td>
<td>Std. Dev.</td>
<td>$0.0111$</td>
<td>$0.0095$</td>
<td>$0.0266$</td>
<td>$0.0232$</td>
</tr>
<tr>
<td>3</td>
<td>Mean</td>
<td>$-1.4813$</td>
<td>$0.7831$</td>
<td>$1.0232$</td>
<td>$-1.2670$</td>
</tr>
<tr>
<td></td>
<td>Std. Dev.</td>
<td>$0.0130$</td>
<td>$0.0172$</td>
<td>$0.0471$</td>
<td>$0.0680$</td>
</tr>
<tr>
<td>4</td>
<td>Mean</td>
<td>$-1.4414$</td>
<td>$0.7565$</td>
<td>$0.9975$</td>
<td>$-1.2300$</td>
</tr>
<tr>
<td></td>
<td>Std. Dev.</td>
<td>$0.2533$</td>
<td>$0.1559$</td>
<td>$0.0210$</td>
<td>$0.2635$</td>
</tr>
</tbody>
</table>

Method 1: The generalized $l_p$ norm estimation algorithm.
Method 2: The generalized $l_p$ ($p = 2$) norm estimation algorithm (GLS).
Method 3: The generalized $l_p$ ($p = 2.5$) norm estimation algorithm.
Method 4: The Bias-correction estimation algorithm.

The SNR is 15 dB in this example. Averaging all the estimated results over 50 simulations, the results are given in Tables V–VII. The accuracy is chosen to be within $\pm 10^{-6}$ and Tables V–VII also include the results of GLS and Bias-correction methods. For the generalized $l_p$ norm estimation algorithm, the averaged estimate values of $p$ are, respectively, 1.99, 1.39, and 1.43 in type (i), (ii), and (iii) disturbances; and the average number of iteration steps are, respectively, 10, 28, and 26 in type (i), (ii), and (iii) disturbances.

The simulation results reveal that the value of $\tilde{p}$ converges to 2 in the Gaussian case, the $l_p$ norm estimators are unbiased for all $p$ and for all noise distributions, and the generalized $l_p$ norm estimation algorithm is more efficient than any other unbiased estimator with the exception of the Gaussian case. Therefore, we are sure that the proposed $l_p$ norm estimation algorithm is a good approach to the parameter estimation problem with unknown measurement noise distribution in nonminimum-phase systems.

VII. CONCLUSION

By modifying the IRLS algorithm via Newton-Raphson method, an unbiased $l_p$ norm estimation algorithm (MIRLS algorithm) has been presented. Based on the MIRLS algorithm, an unbiased generalized $l_p$ norm estimation algorithm is proposed to estimate the unknown parameters of linear time-invariant...
single-input single-output systems with the noisy output variables. Furthermore, if noise is a generalized $p$-Gaussian, the generalized $L_p$ norm parameter estimation algorithm achieves the maximum likelihood estimation. A priori statistical information of the additive noise is assumed to be unknown. How to choose an adequate exponent $p(1 < p < \infty)$ is also introduced. The convergence of the proposed generalized $L_p$ norm parameter estimation algorithm is investigated. For several different types of measurement noise, simulation results illustrate that the unknown parameters can be estimated accurately. Therefore, although a priori information of noise distribution is unknown, the generalized $L_p$ norm algorithm is a good approach to the robust system parameter estimation problem.

**APPENDIX**

In the following, we will consider the convergence problem of $\hat{p}$ for the Gaussian noise case, i.e., does $\hat{p}$ converge to 2 in the Gaussian case? It will be proven that, under the Gaussian noise case, $p$ in (32) tends to 2 with probability one. Fundamentally, we know that if $e(t)$ is a zero mean stationary ergodic random signal, then we have the following results [23]

$$\lim_{N \to \infty} \frac{1}{N} \sum_{t=1}^{N} [e(t) - \bar{e}]^2 \overset{w,p}{\to} E[e(t)]^2$$

$$\lim_{N \to \infty} \frac{1}{N} \sum_{t=1}^{N} [e(t) - \bar{e}]^4 \overset{w,p}{\to} \frac{4}{3} E[e(t)]^4.$$

Thus, we have

$$\lim_{N \to \infty} \hat{\beta}_2 = \lim_{N \to \infty} \frac{1}{N} \sum_{t=1}^{N} [e(t) - \bar{e}]^4 \overset{w,p}{\to} \frac{4}{3} \frac{E[e(t)]^4}{E[e(t)]^2} = \beta_2.$$ 

Furthermore, if $e(t)$ is a Gaussian signal, by (24), we can obtain the following result

$$\beta_2 = \frac{E[e(t)]^4}{\{E[e(t)]^2\}^2} = \frac{3}{2} = 3,$$

i.e.,

$$\lim_{N \to \infty} \hat{\beta}_2 \overset{w,p}{\to} \beta_2 = 3.$$ 

Therefore, if $e(t)$ is a Gaussian signal, we obtain the following result

$$\lim_{N \to \infty} \hat{\beta}_2 \overset{w,p}{\to} \beta_2 = 2,$$

i.e., $\hat{p}$ converges to 2 with probability one in the Gaussian case.

**REFERENCES**


Jeng-Ming Chen received the B.S. degree in electrical engineering from the Tamkang University, Tamsui, Taiwan, R.O.C., and the M.S. degree in materials science and engineering from the National Tsing Hua University, Hsinchu, Taiwan, in 1978 and 1984, respectively. He is presently working toward the Ph.D. degree in the Department of Electrical Engineering, National Tsing Hua University, Hsinchu, Taiwan.

His current research interests include the parameter estimation, signal processing, and estimation theory applications.

Shen-Ching Shern was born in Taiwan, R.O.C. on October 14, 1967. He received the M.S. degree in electrical engineering from the National Tsing Hua University, Hsinchu, Taiwan, R.O.C. in 1991.

He is now a Design Engineer at the Industrial Technology Research Institute, Hsinchu, Taiwan. His main research interests are in the areas of digital signal processing and digital system design.