

# Maximum Likelihood Parameter Estimation of F-ARIMA Processes Using the Genetic Algorithm in the Frequency Domain

Bor-Sen Chen, *Fellow, IEEE*, Bore-Kuen Lee, *Member, IEEE*, and Sen-Chueh Peng, *Member, IEEE*

**Abstract**—This work aims to treat the parameter estimation problem for fractional-integrated autoregressive moving average (F-ARIMA) processes under external noise. Unlike the conventional approaches from the perspective of the time domain, a maximum likelihood (ML) method is developed in the frequency domain since the power spectrum of an F-ARIMA process is in a very explicit and more simple form. However, maximization of the likelihood function is a highly nonlinear estimation problem. Conventional searching algorithms are likely to converge to local maxima under this situation. Since the genetic algorithm (GA) tends to find the globally optimal solution without being trapped at local maxima, an estimation scheme based on the GA is therefore developed to solve the ML parameter estimation problem for F-ARIMA processes from the frequency domain perspective. In the parameter estimation procedure, stability of the F-ARIMA model is ensured, and convergence to the global optimum of the likelihood function is also guaranteed. Finally, several simulation examples are presented to illustrate the proposed estimation algorithm and exhibit its performance.

**Index Terms**—F-ARIMA processes, frequency domain maximum likelihood parameter estimation, genetic algorithm.

## I. INTRODUCTION

RECENTLY, fractal signals, i.e., fractionally differenced noises, have attracted much consideration in signal processing, image processing, geophysical data, network traffic, and computer vision due to the wide variety of data for which they are inherently well suited [2]–[4], [6], [18], [19], [25]–[33]. These processes provide good models for self-similarity and long-range correlation structure observed in several signal processes. Fractal signals are increasingly important candidates for data modeling in a variety of signal processing applications. In contrast to the well-known family of autoregressive moving average (ARMA) processes, fractal signals are characterized by self-similarity and long-range correlation structure. The

ARMA processes provide poor representations for such signals. A fractal signal exhibits a strong long-range correlation with a  $1/f$ -type averaged spectral behavior or, more generally, with a spectral density that is approximately proportional to  $f^{-\gamma}$  with a fractional number  $\gamma > 0$ . In contrast to ARMA processes that are characterized by correlation functions that decay exponentially, fractal signals exhibit the correlation functions that decrease hyperbolically fast [5]. A canonical model of the fractal signal is the fractional Brownian motion introduced by Mandelbrot and Ness [6]. The fractional Brownian motion (fBm) is considered as the  $(H - (1/2))$ th fractional integral of the Brownian motion. The discrete-time version of the fractal signal has also been defined and discussed by Granger and Joyeux [23] and Hosking [24].

Even fractal signals are more suitable for modeling processes with long-range dependence; however, they cannot efficiently model those with both short-range and long-range dependence. In this situation, the F-ARIMA processes, which are modeled as passing a fractal signal through an ARMA filter, are introduced to model the processes with both short-range and long-range dependence [5], [24], [26], [28]. The problem of restoration of a F-ARIMA process after passing through a linear filter (channel) was solved by Chen and Lin [1] by developing a multiscale Wiener filterbank to restore the original signal using the wavelet filterbank technique. However, it was assumed in [1] that the parameters of the F-ARIMA process and additive noise are all known. In practical signal processing, however, these parameters need to be precisely estimated before restoration. This paper is therefore a further extension of the previous work [1] that deals with the estimation problem in F-ARIMA processes.

A parameter estimation algorithm is developed in this study to estimate the parameters of a F-ARIMA process and noise from the received noisy signal. A maximum likelihood (ML) parameter estimation method for fractal signals has been proposed by Wornell and Oppenheim [7] from the time-scale domain perspective, based on the wavelet transform and the expectation-maximization (EM) technique. The ML estimation problem of the parameter for a discrete fractionally differenced Gaussian noise process has also been discussed in [25]. In the above studies, however, the effect of the ARMA filter was not considered. This effect is very important in practical signal processing, e.g., reflections in seismic data processing, intersymbol interference (ISI) in equalization, blur in image processing, etc. The ML parameter estimation problem of F-ARIMA processes has been discussed in the time domain in [5], wherein a modified EM algorithm was applied to solve

Manuscript received January 17, 2000; revised May 24, 2002. This work was supported by the National Science Council under Contracts NSC 90-2213-E-007-019, NSC-90-2213-E-216-009, and NSC90-2213-E-150-001. The associate editor coordinating the review of this paper and approving it for publication was Editor-in-Chief Dr. Arye Nehorai.

B.-S. Chen is with the Department of Electrical Engineering, National Tsing-Hua University, Hsinchu, Taiwan, R.O.C. (e-mail: bschen@moti.ee.nthu.edu.tw).

B.-K. Lee is with the Department of Electrical Engineering, Chung-Hua University, Hsinchu, Taiwan, R.O.C. (e-mail: bklee@chu.edu.tw).

S.-P. Peng is with the Department of Electrical Engineering, National Huwei Institute of Technology, Huwei, Yunlin, Taiwan, R.O.C. (e-mail: speng@nhit.edu.tw).

Publisher Item Identifier 10.1109/TSP.2002.801918.

this problem. In general, the EM method is an initial-condition dependent algorithm, which is likely to converge to a local maximum of the log-likelihood function if the initial condition is not chosen appropriately.

The drawback of the ML parameter estimation methods in the time domain is that they transform explicit parameters of a F-ARIMA process in the frequency domain into implicit parameters of an autocorrelation function in the time domain but in a more complicated form. This transformation increases computational complexity, number of local maxima, and estimation error. Furthermore, the stability of the F-ARIMA process cannot be guaranteed by using the EM algorithm to treat the addressed parameter estimation problem. This drawback will cause the obtained result to be not efficacious. In this study, the parameters of a F-ARIMA process and the additive noise are all estimated directly in the frequency domain from the received signal. The F-ARIMA process is expressed in the frequency domain because the spectral density of a F-ARIMA process has a simpler form. Then, an estimation algorithm, with which the stability of the F-ARIMA model is ensured, is developed from the perspective of the frequency domain, based on the genetic algorithm (GA) to achieve the global optimum of the highly nonlinear likelihood function in the parameter estimation process.

GAs are optimization, machine learning algorithms that were initially inspired from the process of natural selection and evolution of genetics. Unlike the steepest descent approaches to parameter identification and filter design, the GA requires no calculation of gradient and is not susceptible to local maxima that arise from multimodal maximization problems. Therefore, it is more suitable for solving the ML parameter estimation problem for F-ARIMA processes.

In order to treat the parameter estimation problem by the GA, a fitness function must first be formulated, according to the likelihood function. The proposed algorithm begins with a collection of parameter estimates (chromosomes), and each one is evaluated for its fitness in solving the given optimization task. In each generation, the chromosomes with higher fitness values are allowed to mate and bear offspring. These children (new parameter estimates) then form the basis for the next generation. Because of the use of crossover and mutation, this parameter estimation algorithm tends to find the global optimum solution without being trapped at local maxima. The GA was first introduced by Holland [8] and then extensively explored by Goldberg [9]. It has been successfully applied to a variety of optimization problems, such as image processing [10], system identification [11], [12], and fuzzy logic controller design [13]. The GA is employed in the estimation algorithm proposed here to search for the global maximum in the parameter space of the likelihood function of F-ARIMA processes in frequency domain. The concepts presented in [14] and [15] are adopted and modified in this study to guarantee the convergence in probability of the proposed parameter estimation algorithm.

The rest of this paper is organized as follows. The parameter estimation problem is described in Section II. The likelihood function for parameter estimation is formulated in the frequency domain in Section III. The parameter estimation method based on the GA is described in Section IV. Convergence of the proposed algorithm is also discussed. Several simulation results

are presented in Section V, and conclusions are summarized in Section VI.

## II. PROBLEM STATEMENT

The fractional Brownian motion (fBm) was first introduced by Mandelbrot and Ness [6] to formalize the family of signals with  $1/f$  type spectra. The fBm is a generalization of the standard Brownian motion and is defined as

$$B_H(0) = 0$$

$$B_H(t) - B_H(0) = \frac{1}{\Gamma(H+0.5)} \left\{ \int_0^t (t-s)^{H-0.5} dB(s) + \int_{-\infty}^0 [(t-s)^{H-0.5} - (-s)^{H-0.5}] dB(s) \right\}$$

where  $H$  is a parameter with magnitude between 0 and 1, and  $B(s)$  is the standard Brownian motion. In case of  $H = 0.5$ , fBm becomes the standard Brownian motion. The parameter  $H$  is related to the fractal dimension  $D$  of the graph of  $B_H(t)$  by  $D = 2 - H$ . Note that fBm may also be viewed as the  $(H + 0.5)$ th integral of a white noise.

Similar to the continuous-time case, a discrete-time fractal process called discrete fractionally differenced Gaussian noise (fdGn) has been defined by Granger and Joyeux [23] and Hosking [24] through

$$u(n) = (1 - q^{-1})^{-d} v(n)$$

$$= \sum_{k=0}^{\infty} \binom{-d}{k} (-1)^k v(n-k)$$

$$= \sum_{k=0}^{\infty} \frac{(k+d-1)!}{k!(d-1)!} v(n-k)$$

$$= \sum_{k=0}^{\infty} c(k, d) v(n-k)$$

where  $q^{-1}$  is the delay operator with  $q^{-1}v(n) = v(n-1)$ ,  $v(n)$  is a zero-mean, Gaussian, white noise with variance  $\sigma^2$ , and the coefficients  $c(k, d)$ s are the Cesaro numbers. The term  $(1 - q^{-1})^d$  acts as the fractional differencing operation for a fractional number  $d$ . In this paper, we assume that  $0 < d < 0.5$ . The relation between the fractional number  $d$  and the Hurst parameter  $H$  in long memory process in the discrete-time case is  $H = d + 0.5$ .

In [24], Hosking has proven that the process  $u(n)$  is stationary and invertible, i.e.,  $(1 - q^{-1})^d u(n) = v(n)$ , if  $0 < d < 0.5$ . Moreover, since  $v(n)$  is a zero-mean process, so is  $u(n)$ . For  $0 < d < 0.5$ , the power spectrum of a discrete fractal signal  $u(n)$  can be represented as [24]

$$S(\omega) = \frac{2^{-2d} \sigma^2}{(\sin(\omega/2))^{2d}}. \quad (1)$$

Note that  $S(\omega) \approx \sigma^2 \omega^{-2d}$  as  $\omega \rightarrow 0$ . Therefore, the spectrum of the discrete fractal signal has a similar behavior as that of the fBm at low frequencies.

A fractal signal with a power spectrum as in (1) has an autocorrelation asymptotically proportional to  $\tau^{-(1-2d)}$ , where  $\tau$

is the time lag. Therefore, it is suitable for modeling a signal with long-range correlation structure. However, it is not flexible enough to model a signal with short-range correlation structure. Contrarily, an ARMA model with an exponentially decayed correlation function is suitable for modeling short-range correlation of signals but cannot capture their long-range correlation characteristics. In [20], [22], [24], and [26], it is pointed out that a hybrid structure of an input fractal signal and an ARMA filter can be used to capture both short- and long-range dependence in signals. Such a hybrid structure, i.e., the F-ARIMA process, has been found useful in several areas of signal processing in practical applications [1], [25]–[28]. On the other hand, the signal is usually interfered by environment, which can be represented by a measurement noise  $w(n)$ . Consequently, the observed signal  $y(n)$  of the F-ARIMA process can be represented as

$$y(n) = \frac{B(q)}{A(q)} u(n) + w(n) \quad (2)$$

where  $u(n)$  is the input fractal signal with power spectrum as in (1),  $w(n)$  is a zero-mean, Gaussian, white noise with variance  $\sigma_w^2$ , and the rational filter (ARMA channel)  $B(q)/A(q)$  is of the form

$$\frac{B(q)}{A(q)} = \frac{1 + b_1 q^{-1} + \dots + b_{n_b} q^{-n_b}}{1 + a_1 q^{-1} + \dots + a_{n_a} q^{-n_a}}. \quad (3)$$

Assume that the orders  $n_a$  and  $n_b$  are known and that the white noises  $v(n)$  and  $w(n)$  are uncorrelated. Without loss of generality, both the leading coefficient of  $B(q)$  and  $A(q)$  can be set to unity for the purpose of normalization. That is because the gain of  $B(q)/A(q)$  can be absorbed by the parameter  $\sigma$  in (1). For parameter estimation of the F-ARIMA model, the following assumptions are made.

A1) The polynomials  $A(q)$  and  $B(q)$  are coprime.

A2)  $A(q)$  and  $B(q)$  are stable [i.e., all roots of  $A(z)$  and  $B(z)$  are in  $\{z \mid |z| \leq 1\}$ ].

In practical applications, the parameters  $a_1, \dots, a_{n_a}, b_1, \dots, b_{n_b}, d, \sigma, \sigma_w$  of  $B(q)/A(q), u(n)$ , and  $w(n)$  need to be estimated from the noisy output signal  $y(n)$  before designing the wavelet filterbank for restoration of the fractal signals [1]. Unlike the conventional treatment by using the ML method with wavelet multiscale representation [7] in time-scale domain or the EM-algorithm [5], [20] in the time domain, the parameter estimation problem in the processing of the F-ARIMA process in (2) is solved in the frequency domain in the present study.

Let  $\theta$  denote the vector of all the unknown parameters as

$$\theta = [d \ \sigma \ \sigma_w \ a_1 \ \dots \ a_{n_a} \ b_1 \ \dots \ b_{n_b}]^T.$$

The design problem now involves the estimation of the parameter vector  $\theta$  from the  $N$ -points observation  $\{y(n)\}$  for  $n = 0, 1, \dots, N-1$  of a F-ARIMA process corrupted with the external noise in (2).

*Remark 1:* In case of  $B(q)/A(q) = 1$ , the parameter vector  $\theta$  becomes

$$\theta = [d \ \sigma \ \sigma_w]^T.$$

It is then reduced to a parameter estimation problem for fractal signals.  $\square$

### III. LIKELIHOOD FUNCTION FOR PARAMETER ESTIMATION IN THE FREQUENCY DOMAIN

By observing the spectral density  $S(\omega)$  of the fractal signal in (1), it is found that  $S(\omega)$  has a very simple structure. Moreover, the parameters  $d$  and  $\sigma$  of the fractal signal are both in an explicit form in  $S(\omega)$ . Therefore, it is more convenient to estimate the parameters in the frequency domain than in the time domain.

Taking the  $N$ -point discrete Fourier transformation of (2) yields the frequency-domain model as [22], for  $k = 1, \dots, N/2 - 1$

$$Y(e^{j\omega_k}) = \frac{B(e^{j\omega_k})}{A(e^{j\omega_k})} U(e^{j\omega_k}) + W(e^{j\omega_k}), \quad \omega_k = \frac{2\pi k}{N} \quad (4)$$

with

$$Y(e^{j\omega_k}) = \sum_{n=0}^{N-1} y(n) e^{-j\omega_k n} \quad (5)$$

where  $Y(e^{j\omega_k}), U(e^{j\omega_k}),$  and  $W(e^{j\omega_k})$  denote the Fourier coefficients of  $y(n), u(n),$  and  $w(n)$ , respectively, and  $N$  denotes the number of data points. The zero frequency is not considered in (4) due to the singularity of  $S(\omega)$ , as shown in (1). In order to simply use the fast Fourier transform (FFT),  $N$  is chosen as  $2^l$  for some positive integer  $l$ . Technically, the FFT can be applied to the case that  $N$  is a product of powers of small prime numbers.

To derive the likelihood function for parameter estimation in the frequency domain, we need the following results.

*Theorem 1:* With the Gaussian assumptions of the uncorrelated white noises  $v(n)$  and  $w(n)$  and the stability assumption A2) on  $A(q)$  and  $B(q)$ , the Fourier coefficients  $Y(e^{j\omega_k})$  at frequencies  $\omega_k, k = 1, \dots, N/2 - 1$  are approximately statistically independent complex Gaussian random variables with probability densities as

$$p(Y(e^{j\omega_k})) \approx \frac{1}{\pi \text{var}(Y(e^{j\omega_k}))} \exp\left(-\frac{Y(e^{j\omega_k})Y(e^{-j\omega_k})}{\text{var}(Y(e^{j\omega_k}))}\right) \quad (6)$$

where  $\text{var}(Y(e^{j\omega_k}))$ , which is the data variance at frequency  $\omega_k$  as  $N \rightarrow \infty$ , is given by

$$\text{var}(Y(e^{j\omega_k})) = \frac{B(e^{j\omega_k})B(e^{-j\omega_k})}{A(e^{j\omega_k})A(e^{-j\omega_k})} NS(\omega_k) + N\sigma_w^2 \quad (7)$$

where  $S(\omega_k)$  and  $\sigma_w^2$  are the power spectra of  $u(n)$  and  $w(n)$ , respectively.

*Proof:* Before presenting the proof of (6), the F-ARIMA process in (2) is rewritten as

$$y(n) = x(n) + w(n) \quad (8)$$

$$x(n) = \frac{B(q)}{A(q)} \frac{1}{(1 - q^{-1})^d} v(n). \quad (9)$$

Recall that  $v(n)$  is the white Gaussian driving noise with zero mean and variance  $\sigma^2$ .

Let us denote the joint  $k$ th-order cumulants [22] of  $w(n)$  and  $v(n)$  as  $C_k^w(n_1, n_2, \dots, n_{k-1})$  and  $C_k^v(n_1, n_2, \dots, n_{k-1})$  at times  $n, n + n_1, \dots, n + n_{k-1}$ , respectively. Since the white

noises  $w(n)$  and  $v(n)$  are zero-mean Gaussian, all the moments of  $w(n)$  and  $v(n)$  are finite, and the following conditions hold:

$$\sum_{n_1, n_2, \dots, n_{k-1} = -\infty}^{\infty} |C_k^w(n_1, n_2, \dots, n_{k-1})| < \infty \quad (10)$$

$$\sum_{n_1, n_2, \dots, n_{k-1} = -\infty}^{\infty} |C_k^v(n_1, n_2, \dots, n_{k-1})| < \infty \quad (11)$$

for any finite  $k$ .

Following from [22, Th. 4.4.1] with the finite cumulant condition (10) on  $w(n)$ , the Fourier coefficients  $W(e^{j\omega_k})$ ,  $k = 1, 2, \dots, N/2 - 1$  are with the asymptotical independent complex Gaussian distribution  $N^c(0, N\sigma_w^2)$  as  $N \rightarrow \infty$ . Note that [22, Th. 4.4.1] holds only for conventional stationary processes. For the F-ARIMA process  $x(n)$ , it is modified in [36]. Under assumption A2), the function  $|B(e^{j\omega})|/|A(e^{j\omega})|$  is a positive and continuous function of  $\omega$ . Then, following from assumption A2) and (11), the Fourier coefficients  $X(e^{j\omega_k})$ ,  $k = 1, 2, \dots, N/2 - 1$  are with the asymptotical independent complex Gaussian distribution  $N^c(0, Nf_x(w_k))$  as  $N \rightarrow \infty$ , where  $f_x(w_k)$  denotes the spectral density of  $x(n)$  as

$$f_x(w_k) = \frac{B(e^{jw_k})B(e^{-jw_k})}{A(e^{jw_k})A(e^{-jw_k})} S(w_k)$$

where  $S(w_k)$  denotes the power spectrum of  $u(n)$  at  $w_k$ .

Since we assume that the white noises  $v(n)$  and  $w(n)$  are uncorrelated, based on the above reasoning,  $Y(e^{j\omega_k}) = X(e^{j\omega_k}) + W(e^{j\omega_k})$  for  $k = 1, 2, \dots, N/2 - 1$  are with the asymptotical independent Gaussian distribution  $N^c(0, \text{var}Y(e^{j\omega_k}))$ . Then, we get the probability distribution of  $Y(e^{j\omega_k})$  as (6). Moreover, by the fact that [22]

$$\begin{aligned} E(U(e^{j\omega_k})U(e^{-j\omega_k})) &= NS(\omega_k) \\ E(W(e^{j\omega_k})W(e^{-j\omega_k})) &= N\sigma_w^2 \end{aligned}$$

one has  $\text{var}Y(e^{j\omega_k}) = N\sigma_w^2 + Nf_x(w_k)$ , and the result in (7) can be easily obtained.  $\square$

Since the data length  $N$  is large enough, we have

$$S(\omega_k) \approx S(\omega)|_{\omega_k=2\pi k/N} = \frac{2^{-2d}\sigma^2}{(\sin(\omega_k/2))^{2d}}$$

The data variance  $\text{var}(Y(e^{j\omega_k}))$  at the frequency  $\omega_k$  is then expressed as

$$\text{var}(Y(e^{j\omega_k})) = N\sigma_w^2 + \left| \frac{1 + b_1 e^{j\omega_k} + \dots + b_{n_b} e^{jn_b \omega_k}}{1 + a_1 e^{j\omega_k} + \dots + a_{n_a} e^{jn_a \omega_k}} \right|^2 \frac{2^{-2d}N\sigma^2}{(\sin(\omega_k/2))^{2d}} \quad (12)$$

where  $|A|$  denotes the absolute value of a complex number  $A$ . Invoking the approximate statistical independence of  $Y(e^{j\omega_k})$ ,  $k = 1, \dots, N/2 - 1$ , the log-likelihood function is

$$\begin{aligned} L_Y(\theta) &= \sum_{k=1}^{N/2-1} \log p(Y(e^{j\omega_k})) \\ &\approx \sum_{k=1}^{N/2-1} \left[ -\log \pi - \log \text{var}(Y(e^{j\omega_k})) \right. \\ &\quad \left. - \frac{|Y(e^{j\omega_k})|^2}{\text{var}(Y(e^{j\omega_k}))} \right]. \end{aligned}$$

The term  $\log \pi$  stays invariant during the maximization procedure and, hence, can be discarded. Consequently, an equivalent representation of the log-likelihood function is obtained as

$$L_Y(\theta) \approx \sum_{k=1}^{N/2-1} \left[ -\log \text{var}(Y(e^{j\omega_k})) - \frac{|Y(e^{j\omega_k})|^2}{\text{var}(Y(e^{j\omega_k}))} \right]. \quad (13)$$

Substituting (12) into (13) yields (14), shown at the bottom of the page.

*Remark 2:* In case of  $B(q)/A(q) = 1$ , the log-likelihood function in (14) is reduced to

$$L_Y(\theta) \approx - \sum_{k=1}^{N/2-1} \left\{ \log \left( \frac{2^{-2d}N\sigma^2}{(\sin(\omega_k/2))^{2d}} + N\sigma_w^2 \right) + \frac{|Y(e^{j\omega_k})|^2}{\left( \frac{2^{-2d}N\sigma^2}{(\sin(\omega_k/2))^{2d}} + N\sigma_w^2 \right)} \right\}. \quad (15)$$

Then, the ML parameter estimation problem of F-ARIMA processes in the frequency domain is to find an optimal parameter vector  $\theta^*$  to solve the following maximization problem:

$$\max_{\theta} L_Y(\theta). \quad (16)$$

$$\begin{aligned} L_Y(\theta) \approx - \sum_{k=1}^{N/2-1} &\left[ \log \left( \left| \frac{1 + b_1 e^{j\omega_k} + \dots + b_{n_b} e^{jn_b \omega_k}}{1 + a_1 e^{j\omega_k} + \dots + a_{n_a} e^{jn_a \omega_k}} \right|^2 \frac{2^{-2d}N\sigma^2}{(\sin(\omega_k/2))^{2d}} + N\sigma_w^2 \right) \right. \\ &\left. + \frac{|Y(e^{j\omega_k})|^2}{\left( \left| \frac{1 + b_1 e^{j\omega_k} + \dots + b_{n_b} e^{jn_b \omega_k}}{1 + a_1 e^{j\omega_k} + \dots + a_{n_a} e^{jn_a \omega_k}} \right|^2 \frac{2^{-2d}N\sigma^2}{(\sin(\omega_k/2))^{2d}} + N\sigma_w^2 \right)} \right] \quad (14) \end{aligned}$$

By observing the log-likelihood function as in (14), it is found that  $L_Y(\theta)$  is a highly nonlinear function of  $\theta$ , especially for the fractal parameter  $d$ . There may exist many local optima in the maximization of the log-likelihood function in (16). By employing the conventional methods to treat this optimization problem, a local solution is usually obtained. Furthermore, some derivative information or the complete knowledge of the problem structure and parameters are also required for conventional optimization algorithms. These drawbacks render them not suitable to treat our problem. Contrarily, the GA tends to find the global solution of the ML problem in (16) without being trapped at local optima and does not require the information of derivative or problem structure. The GA is a parallel and global optimal search technique that copies natural genetic operations to simultaneously evaluate many points in the parameter space and more likely to converge toward the global solution of the optimization problem in (16). Hence, it is more suitable than the other optimization algorithms to treat our optimization problem.

#### IV. PARAMETER ESTIMATION VIA THE GENETIC ALGORITHM

The GA is a stochastic optimization algorithm that was originally motivated by the mechanisms of natural selection and evolution of genetics. The underlying principles of the GA were first proposed by Holland in 1962 [17], whereas the mathematical framework was developed in the late 1960s and was presented in Holland's pioneering book [8]. In the following, a parameter estimation algorithm is developed based on the GA to estimate the parameter vector  $\theta$  of the F-ARIMA process in (2) by carrying out maximization of the log-likelihood function in (16).

##### Selection of Search Space

By using the GA to solve the problem of maximization of the log-likelihood function  $L_Y(\theta)$  in (16), the search space of the parameter vector  $\theta$  must be specified properly beforehand. This is because an appropriate choice of the search space may speed up the convergence of the GA. In general, the search space can be specified based on the characteristics of the parameters. In case of  $B(q)/A(q) \neq 1$ , the choice of the search space is more difficult than that of  $B(q)/A(q) = 1$ . This is because both the stability of the polynomials (i.e., all the roots of  $A(z)$  and  $B(q)$  must be in  $\{z \mid |z| \leq 1\}$  [34]) must be taken into consideration. In general, determining the range of the coefficients of  $A(q)$  and  $B(q)$  to guarantee their stability is difficult work, except for low order systems. In most of the applications, the entire search space  $\Theta$  of the parameter vector  $\theta$  may not simply be described by the rectangular form as

$$\Theta: \{\theta \mid \theta_i^L \leq \theta_i \leq \theta_i^U, \text{ for } i = 1, \dots, M\}$$

where  $M$  denotes the number of the parameter. Let us consider a second-order ARMA model  $B(z)/A(z)$  with the denominator polynomial  $A(z) = 1 + a_1 z^{-1} + a_2 z^{-2}$ . The stable region  $\Theta_a$  in this case is of the triangular form [see Fig. 1(a)]. It cannot be expressed in the above rectangular form. Under this situation, the proposed GA is no longer applied directly. To overcome this drawback, one can divide the stable region  $\Theta_a$  into

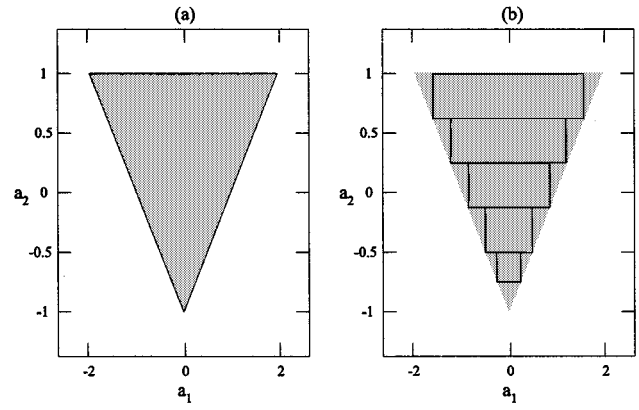


Fig. 1. (a) Stable region for a second-order system. (b) One of the division of the stable region for a second-order system.

several regions  $\Theta_{a_1}, \dots, \Theta_{a_J}$ , which are all stable and of the rectangular form. Fig. 1(b) illustrates one of the division for the second-order system. The same conclusion is also applied to polynomial  $B(z)$ . Therefore, the entire search space  $\Theta$  can be divided into a union of regions  $\Theta_1, \dots, \Theta_J$  for some positive integer  $J$ . Now, by applying the proposed GA to each region  $\Theta_i$  to obtain the optimal estimates  $\theta_i^*$ , for  $i = 1, 2, \dots, J$ , the global or near global estimate  $\theta^*$  can be derived by

$$\theta^* = \arg \max_{\theta \in \{\Theta_1, \Theta_2, \dots, \Theta_J\}} L_Y(\theta). \quad (17)$$

The GA searches for the ML parameter only inside the stable parameter space, and the stability of the  $A(z)$  and  $B(z)$  is guaranteed. This leads to a remarkable reduction of search space and saves much effort of computation.

*Remark 3:* For the case in which  $n_a = 2$ , the triangular stable region in Fig. 1(a) can be transformed into a rectangle in a new coordinate system. However, this approach is hard to apply to high-degree cases.  $\square$

In GA, the parameter vector  $\theta$  to be searched to solve the ML parameter estimation problem in (16) is represented by a population of binary strings. The choice of the bit number for each parameter depends on the desired "resolution" we want in the search space. For simplicity, take searching in region  $\Theta_i$ , for example. With binary coding, the resolution of each parameter can be calculated as

$$R_k = \frac{\theta_k^U - \theta_k^L}{2^{\ell_k} - 1} \quad (18)$$

where

- $R_k$  resolution;
- $\theta_k^U$  upper bound of the search range;
- $\theta_k^L$  lower bound of the search range;
- $\ell_k$  bit number for the  $k$ th parameter in  $\Theta_i$ .

For any  $k$ ,  $1 \leq k \leq M$ , the  $k$ th parameter  $\theta_k$  is chosen as

$$\theta_k = \theta_k^L + \mathbf{x}_k \mathbf{R}_k$$

where  $\mathbf{x}_k$  is coded using  $\ell_k$  bits such that  $\theta_k^L \leq \theta_k \leq \theta_k^U$  holds.

##### Fitness Function

The degree of fitness depends on the performance of the possible solution represented by that particular string. The larger the

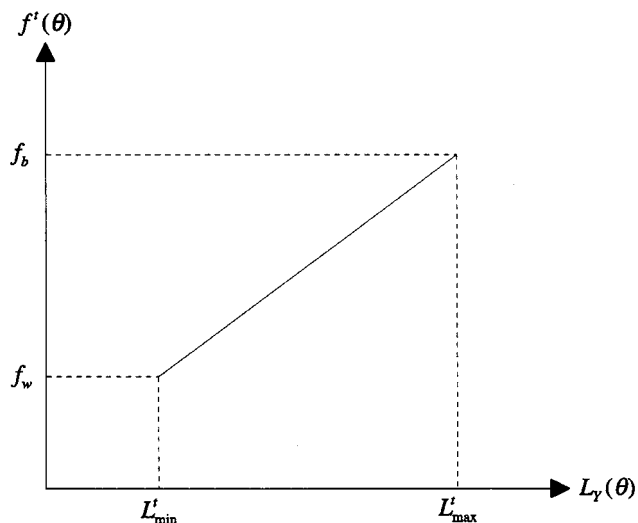


Fig. 2. Relation between  $f^t(\theta)$  and  $L_Y(\theta)$ .

likelihood, the higher the fitness. There are a number of methods to perform this mapping, which are known as fitness techniques. The mapping method adopted in the present study is illustrated in Fig. 2. The fitness function  $f^t(\theta)$  in generation  $t$  of a parameter vector  $\theta$  with log-likelihood function  $L_Y(\theta)$  is expressed by

$$f^t(\theta) = f_w + \frac{f_b - f_w}{L_{\max}^t - L_{\min}^t} (L_Y(\theta) - L_{\min}^t) \quad (19)$$

where  $L_{\max}^t$  and  $L_{\min}^t$  denote the maximum and minimum values of the likelihood function  $L_Y(\theta)$  in the population of generation  $t$ , respectively;  $f_b$  and  $f_w$  are the prescribed best and worst fitness values, respectively. Since  $f^t(\theta)$  is linearly proportional to  $L_Y(\theta)$ , the  $\theta^*$ , which maximizes  $f^t(\theta)$ , also maximizes  $L_Y(\theta)$ . Therefore, the ML parameter estimation problem in (16) is equivalent to the following fitness optimization problem:

$$\max_{\theta} f^t(\theta) \quad (20)$$

Our parameter estimation problem is to develop a GA to solve the above fitness optimization problem, which is described by the following genetic operators.

### Genetic Operators

The most important and basic operations for the GA for solving (20) are maintaining, reproduction, crossover, and mutation. A brief description of these operations is presented in the following. For more detailed introduction, see the fundamental textbooks [8] and [9].

- **Maintaining:** This is a process of copying the best string in this generation (with the highest fitness) to the next generation. The purpose of this operation is to ensure that the best string in the next generation is at least no worse than the best one in this generation.
- **Reproduction:** Reproduction is a process in which individual strings are copied and put in a mating pool for further genetic operations according to their fitness values.

The probability of the  $i$ th string (corresponding to the  $i$ th parameter vector  $\theta_i$ ) with fitness value  $f^t(\theta_i)$  reproduced for mating in the next generation is

$$P_i^t = \frac{f^t(\theta_i)}{\sum_{j=1}^{n_p} f^t(\theta_j)} \quad (21)$$

where  $n_p$  is the population size specified by the designer.

- **Crossover:** Crossover provides a mechanism for exchanging information in two strings via probabilistic decision. Combined with reproduction, it is an effective way of exchanging information and combining portions of high-quality solutions.
- **Mutation:** Mutation is occasional alteration of each bit of a chromosome from 0 to 1 or from 1 to 0 with a small probability  $P_m$ . The purpose of mutation is to introduce occasional perturbation to the estimated parameters to ensure that all points in the search space can ultimately be reached.

The algorithm begins with a population of randomly generated chromosomes. Each chromosome is decoded into the corresponding parameter vector and evaluated for its fitness value in solving the optimization problem in (20). At each generation, chromosomes mate and bear offspring. Note that the best candidate in one generation is retained in the next generation to ensure that the best candidate in the next generation is at least as good as the present one.

An easy way to search the entire parameter space  $\Theta$  is to independently apply the GA to each individual rectangular region  $\Theta_i$ ,  $1 \leq i \leq J$ . Then, the ultimate global or near-global estimate  $\theta^*$  can be derived according to (17). The steps of the GA-based parameter estimation algorithm in a region  $\Theta_i$  are listed as follows.

### Genetic-Based Parameter Estimation Algorithm:

- Step 0)** Select the parameter space  $\Theta_i$  properly for the parameter vector  $\theta$ .
- Step 1)** Randomly generate a population of binary strings in  $\Theta_i$  (generation  $t=0$ ).
- Step 2)** Decode each string into the corresponding parameter vector.
- Step 3)** Calculate the likelihood function of each parameter vector according to (14).
- Step 4)** Calculate the fitness values according to (19).
- Step 5)** Perform the basic operations of the GA, i.e., maintaining, reproduction, crossover, and mutation.
- Step 6)** Increase the generation index  $t$  by 1. If the stopping criterion is not satisfied, go to **Step 2**; otherwise, stop the algorithm.

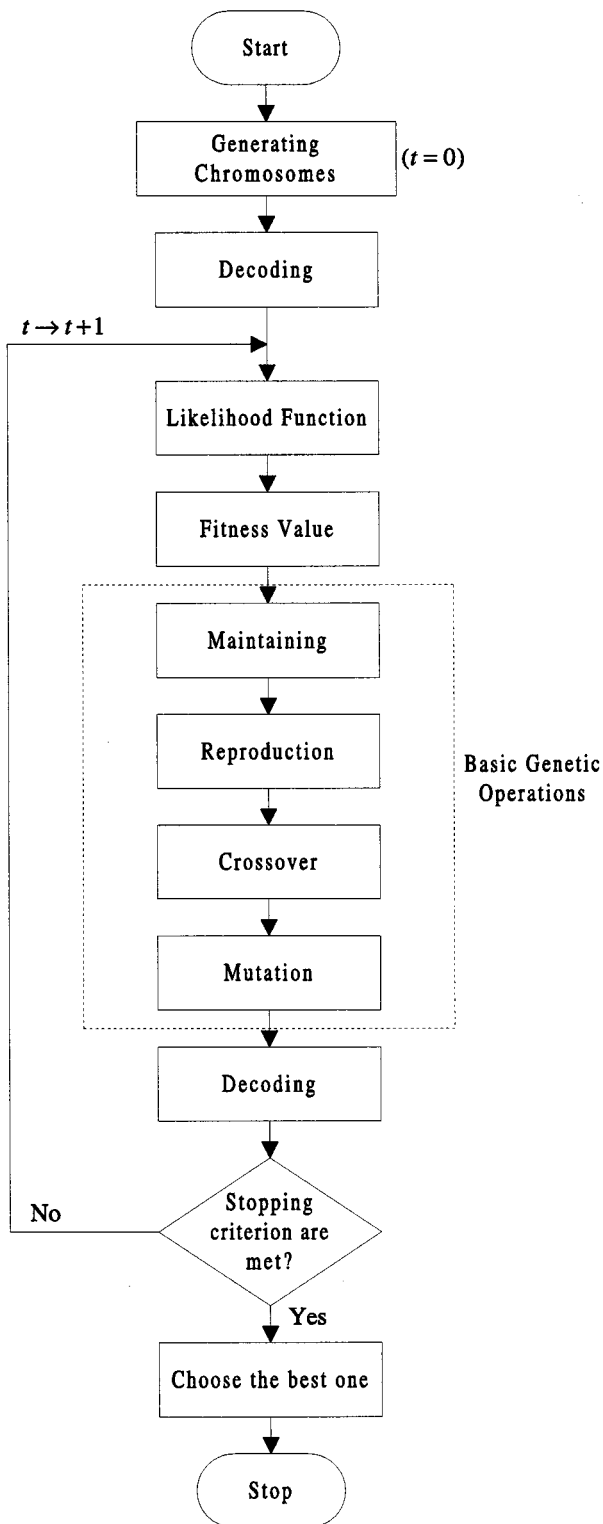


Fig. 3. Flowchart of the GA-based parameter estimation algorithm.

The stopping criterion can be specified as the number of generations tolerated for no improvement on the value of the likelihood function and/or the maximum number of total generations to be performed. The decoded parameter vector corresponding to the best chromosome at the last generation is the solution to the problem of ML parameter estimation in (20). The whole algorithm is summarized in Fig. 3.

### Convergence Discussion of the Parameter Estimation Algorithm

Convergence of the GA has been analyzed in some recent studies [14], [15]. The concept presented in [15] is adopted here to prove the convergence of the proposed parameter estimation algorithm to the global optimum  $\theta^*$  of  $L_Y(\theta)$  [or  $f^t(\theta)$ ].

*Theorem 2:* The estimated parameter converges to an optimal estimate  $\theta^*$  that attains the global maximum of the log-likelihood function  $L_Y(\theta)$  in (14).

*Proof:* First, note that the GA is independently applied to each individual rectangular region  $\Theta_i$ ,  $1 \leq i \leq J$ . To obtain the claimed result, it is sufficient to prove that in any fixed region  $\Theta_i$ , the estimated parameter converges to an optimal estimate that attains the global maximum of the log-likelihood function  $L_Y(\theta)$  in  $\Theta_i$ . Therefore, we will only discuss the problem in a single rectangular region in the following.

Since each chromosome consists of only a finite number of bits, the estimated parameters represented by these chromosomes are actually quantized values. Let  $L_{\max}$  denote the global maximum of the log-likelihood function (14) or (15), and let  $L_{\max}^Q$  denote the maximum of the log-likelihood function evaluated for all possible chromosomes. The case wherein the quantization error in each estimated parameter is zero is called ideal matching. In this case,  $L_{\max}^Q$  is equal to  $L_{\max}$ . The expected value of the log-likelihood function is shown in the following to converge to  $L_{\max}^Q$  at a specific rate.

Let us examine the generation of a child chromosome in generation  $(t+1)$  from its parent chromosomes in generation  $t$ . First, the parent chromosomes  $\mathbf{x}_i$  and  $\mathbf{x}_j$  are selected from the population  $s$  according to the probability in (21). Next, a uniformly distributed crossover site  $c_s \in \{1, 2, \dots, \ell-1\}$  is chosen, and the crossover operation is performed. We denote the probability of a child chromosome generated from parents  $\mathbf{x}_i$  and  $\mathbf{x}_j$  with a crossover site  $c_s$  by  $P_{ijc_s}$ . Then, each chromosome thus generated undergoes the process of mutation. Let us denote the mutation pattern by  $u$  and the probability of occurrence of this pattern by  $P_u$ . Since the mutation operation is performed independently at each bit of a chromosome, the set  $U$  consisting of all possible mutation patterns is of size  $2^\ell$ . The likelihood function corresponding to the child chromosome in generation  $(t+1)$ , which is generated from parents  $\mathbf{x}_i$  and  $\mathbf{x}_j$  with crossover site  $c_s$  and mutation pattern  $u$ , is represented by  $L_{ijc_s u}^{t+1}$ .

Some notations that are useful in the analysis of convergence property are introduced in the following:

- $\delta_{ij}$  difference in the likelihood functions of any two chromosomes  $\mathbf{x}_i$  and  $\mathbf{x}_j$  with corresponding likelihood functions  $L_i$  and  $L_j$ , respectively, i.e.,  $\delta_{ij} = |L_i - L_j|$ ;
- $\bar{\delta}$  minimum of  $\delta_{ij}$  subject to  $\delta_{ij} \neq 0$ ;
- $n_c^t$  number of all possible chromosomes with likelihood function larger than  $L_{\max}^t$ ;
- $S$  set of all possible  $s$  (remember that  $s$  represents the population of chromosomes).

The chromosome with likelihood function  $L_{\max}^t$  in generation  $t$  is copied directly in the next generation, whereas, all other chromosomes in generation  $(t+1)$  are generated from the current population  $s$  by the three operations, i.e., reproduction, crossover, and mutation. Assume that the maximum value

of the likelihood functions of all the newly generated chromosomes (excluding the one copied directly) is  $L_{ijc_s u}^{t+1}$ . The expected value of  $L_{\max}^{t+1}$  conditioned on population  $s$  at generation  $t$  is, therefore

$$\begin{aligned} E(L_{\max}^{t+1}|s) &= \sum_{i,j \in S} \sum_{c_s} \sum_{u \in U} P_{ijc_s} P_u \max(L_{\max}^t, L_{ijc_s u}^t) \\ &= \sum_{i,j,c_s,u} P_{ijc_s} P_u \max(L_{\max}^t, L_{ijc_s u}^t). \end{aligned} \quad (22)$$

Define a notation  $\delta_{ijc_s u}^t$  as

$$\delta_{ijc_s u}^t \equiv L_{ijc_s u}^{t+1} - L_{\max}^t, \quad \forall L_{ijc_s u}^{t+1} > L_{\max}^t.$$

Then, (22) can be expressed as

$$\begin{aligned} E(L_{\max}^{t+1}|s) &= \sum_{L_{ijc_s u}^{t+1} \leq L_{\max}^t} P_{ijc_s} P_u L_{\max}^t \\ &\quad + \sum_{L_{ijc_s u}^{t+1} > L_{\max}^t} P_{ijc_s} P_u (L_{\max}^t + \delta_{ijc_s u}^t) \\ &= \sum_{i,j,c_s,u} P_{ijc_s} P_u L_{\max}^t + \sum_{L_{ijc_s u}^{t+1} > L_{\max}^t} P_{ijc_s} P_u \delta_{ijc_s u}^t \\ &= L_{\max}^t \sum_{i,j,c_s,u} P_{ijc_s} P_u + \sum_{L_{ijc_s u}^{t+1} > L_{\max}^t} P_{ijc_s} P_u \delta_{ijc_s u}^t \\ &= L_{\max}^t + \sum_{L_{ijc_s u}^{t+1} > L_{\max}^t} P_{ijc_s} P_u \delta_{ijc_s u}^t. \end{aligned} \quad (23)$$

Each  $\Sigma$  in (23) is performed for all  $i, j, c_s$ , and  $u$  subject to the constraint stated previously. The number of terms contained in the summation of the last equality in (23) is  $n_c^t$ . If  $L_{\max}^t < L_{\max}^Q$ , i.e., the best chromosome has not been reached,  $n_c^t$  is a positive integer. In practical applications, mutation rate  $P_m \ll 0.5$ . Therefore, the minimum value  $\bar{P}$  of  $P_u$  for all  $u \in U$  is

$$\bar{P} = (P_m)^\ell$$

i.e., the probability that all bits are mutated.

Substitute the previous relation into (23)

$$\begin{aligned} E(L_{\max}^{t+1}|s) &\geq L_{\max}^t + P_{\min} \bar{P} \sum_{L_{ijc_s u}^{t+1} > L_{\max}^t} \min(\delta_{ijc_s u}^t) \\ &\geq L_{\max}^t + n_c^t P_{\min} \bar{P} \min(\delta_{ijc_s u}^t) \\ &\geq L_{\max}^t + n_c^t P_{\min} \bar{P} \bar{\delta} \end{aligned}$$

where  $P_{\min}$  is the minimum of all possible  $P_{ijc_s}$ . The expected value of  $L_{\max}^{t+1}$  is then obtained as

$$\begin{aligned} E(L_{\max}^{t+1}) &= E(E(L_{\max}^{t+1}|s)) \\ &= \sum_{s \in S} P_s E(L_{\max}^{t+1}|s) \\ &\geq \sum_{s \in S} P_s L_{\max}^t + \sum_{s \in S} P_s P_{\min} \bar{P} \bar{\delta} n_c^t \\ &= E(L_{\max}^t) + P_{\min} \bar{P} \bar{\delta} E(n_c^t). \end{aligned} \quad (24)$$

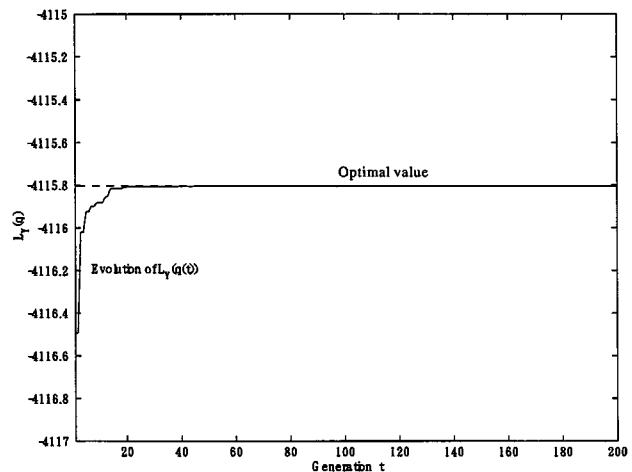


Fig. 4. Convergence of the average  $L_Y(\theta)$  versus generation in Example 1.

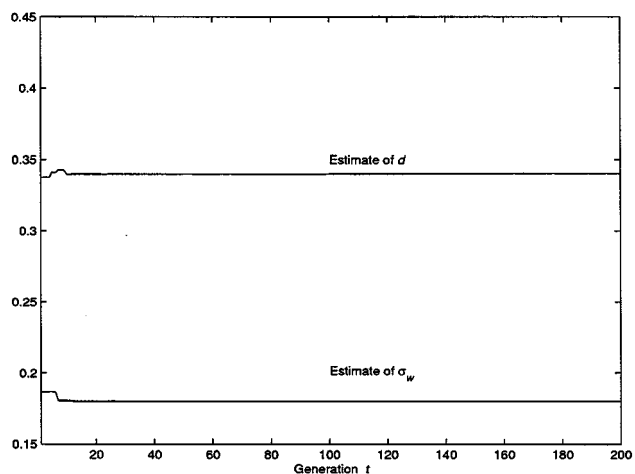


Fig. 5. Typical trace of parameter estimates in Example 1.

Equation (24) implies that

$$E(L_{\max}^t) \geq E(L_{\max}^0) + P_{\min} \bar{P} \sum_{k=0}^{t-1} E(n_c^k).$$

Equation (24) illustrates that the expected value  $E(L_{\max}^{t+1})$  is larger than  $E(L_{\max}^t)$  at least by the value  $P_{\min} \bar{P} \bar{\delta} E(n_c^t)$ . This value is positive, provided that  $E(L_{\max}^t)$  is smaller than  $L_{\max}^Q$ . On the other hand, it is zero, provided that  $E(L_{\max}^t)$  equals  $L_{\max}^Q$ . Consequently,  $\lim_{t \rightarrow \infty} E(L_{\max}^t) = L_{\max}^Q$ . Therefore, in each rectangular region, the estimated parameter converges to an optimal estimate that attains the global maximum of the log-likelihood function  $L_Y(\theta)$  in (14).  $\square$ .

*Remark 4:* The measure of the set of parameter vectors, in which the  $A(z)$  and  $B(z)$  have common factors, is zero with respect to the entire parameter search space. Therefore, with probability 1, the estimated versions of  $A(q)$  and  $B(q)$  are coprime in any generation. On the other hand, by searching in the preselected parameter space, the parameter estimates obey assumption A2). The claim of the above theorem will be verified by several simulation examples in Section V. Since the likelihood function  $L_Y(\theta)$  in frequency domain only considers the amplitude  $|A(e^{jW_k})/B(e^{jW_k})|$ , assumptions A1) and A2) are



TABLE I  
MEAN ESTIMATES AND THEIR STANDARD DEVIATIONS (STD) OF THE PARAMETER VECTOR  $[d \ \sigma_w] = [0.35 \ 0.2]$  USING THE PROPOSED GA-BASED METHOD AND THE ML-EM METHOD VIA THE MONTE CARLO SIMULATION WITH 100 RUNS IN EXAMPLE 1

Estimation Method	Estimated Parameters	Mean ( $\hat{d}$ , $\hat{\sigma}_w$ )	STD (std $\hat{d}$ , std $\hat{\sigma}_w$ )
The proposed GA-based method		(0.3429, 0.1929)	(0.1136, 0.0678)
ML-EM method with initial condition (0.34, 0.18)		(0.3456, 0.1850)	(0.1288, 0.0990)
ML-EM method with initial condition (0.10, 0.01)		(0.3401, 0.0498)	(0.1334, 0.2254)

necessary in order to guarantee the uniqueness of the optimal estimate  $\theta^*$ . We should note that convergence of the estimated parameters to the true value is not provided in the above theorem. Parameter convergence to true value is another significant topic for the GA-based parameter estimation problem for F-ARIMA processes. It needs more effort and will be treated in further research.

## V. NUMERICAL SIMULATIONS

In this section, several numerical simulation examples are presented to illustrate the proposed parameter estimation method and exhibit its performance. Comparisons with the results reported in [5] and [20] are also given to demonstrate the superior performance of the proposed method. For convenience, to compare the proposed method with the ML-EM method in [5] and [20], the parameter  $\sigma$  is set to be unity throughout this section.

*Example 1—A F-ARIMA Process Without ARMA Part:* In this example, the input fractal signal is generated according the formula in [23] and [24] with data length  $N = 1024$  and parameter  $d = 0.35$ . The value  $\sigma_w$  of the additive noise is set to be 0.2. Let  $\theta$  denote the parameter vector as

$$\theta = [d \ \sigma_w]^T = [\theta_1 \ \theta_2]^T = [0.35 \ 0.2]^T.$$

An appropriate choice of the admissible set  $\Theta$  of the parameter vector  $\theta$  can be specified as

$$\Theta: \{\theta \mid 0 < \theta_1 < 0.5, 0 \leq \theta_2 \leq 1\}.$$

In order to obtain an accurate result, the resolution for every parameter is assumed to be slightly finer than  $1.6 \times 10^{-5}$ . According to the discussion in Section IV, we have

$$R_1 = \frac{0.5 - (-0.5)}{2^{\ell_1} - 1} < 1.6 \times 10^{-5}$$

$$R_2 = \frac{1 - 0}{2^{\ell_2} - 1} < 1.6 \times 10^{-5}.$$

The required bit numbers  $\ell_1$  and  $\ell_2$  for  $\theta_1$  and  $\theta_2$  are therefore derived as

$$\ell_1 = 16 > \frac{\log_{10} \left( \frac{1}{1.6 \times 10^{-5}} + 1 \right)}{\log_{10} 2} = 15.9316$$

$$\ell_2 = 16 > \frac{\log_{10} \left( \frac{1}{1.6 \times 10^{-5}} + 1 \right)}{\log_{10} 2} = 15.9316.$$

Consequently, the chromosome of length  $\ell_1 + \ell_2 = 32$  is derived, which leaves the search space  $\Theta$  with  $2^{32}$  alternatives.

By using the proposed GA-based method to treat the parameter estimation problem, the genetic parameters are chosen as

$$\begin{aligned} \text{Population Size } n_p &= 200 \\ \text{Mutation Probability } P_m &= 0.1 \\ \text{Best Fitness Value } f_b &= 100 \\ \text{Worst Fitness Value } f_w &= 10. \end{aligned}$$

After 200 generations, the average estimated parameters are obtained as the following via Monte-Carlo simulation with 100 runs:

$$\hat{d} = 0.3429, \quad \hat{\sigma}_w = 0.1929.$$

Convergence of the mean log-likelihood function  $L_Y(\theta)$  for this case is illustrated in Fig. 4, in which the dash-dotted line indicates the theoretical maximal value of  $L_Y(\theta)$ . Since the algorithm simultaneously searches for many points in the search space  $\Theta$ , the proposed method converges quickly and asymptotically achieves the ML as generation  $t$  increases. Among the 100 runs in the Monte-Carlo simulation, a typical trace of parameter estimates is shown in Fig. 5.

A comparison of the present results with those obtained by the conventional ML-EM algorithm via Monte-Carlo simulation with 100 runs is listed in Table I. It is shown that both the results of the proposed GA-based method and the ML-EM method with a “good” initial condition are almost the same as the values of true parameters. This means that both the proposed GA-based method and the ML-EM algorithm are efficient to treat this ML parameter estimation problem. On the other hand, it is also shown that a worse performance is obtained if the initial condition of the ML-EM method is not chosen properly. This is because the EM algorithm is trapped at local maximum when a bad initial condition is given.

Usually, choice of the data number  $N$  should depend on the decaying rate of the correlation function of the observed signal. A fractal signal has an autocorrelation asymptotically proportional to  $\tau^{-(1-2d)}$ , where  $\tau$  is the time lag. In general, when the fractional number  $d$  is close to 0.5, the correlation decays very slowly. In this situation,  $N$  could be very large. In the example, in order to use the fast Fourier transform (FFT) algorithm, we choose  $N = 2^{10} = 1024$ , and the accuracy of parameter estimation is acceptable.

*Example 2—A F-ARIMA Process With an AR Model:* The input fractal signal is taken to be the same as that in Example 1. The AR model is assumed to be the second-order system

$$\frac{B(z)}{A(z)} = \frac{1}{1 - 0.4z^{-1} + 0.8z^{-2}}.$$

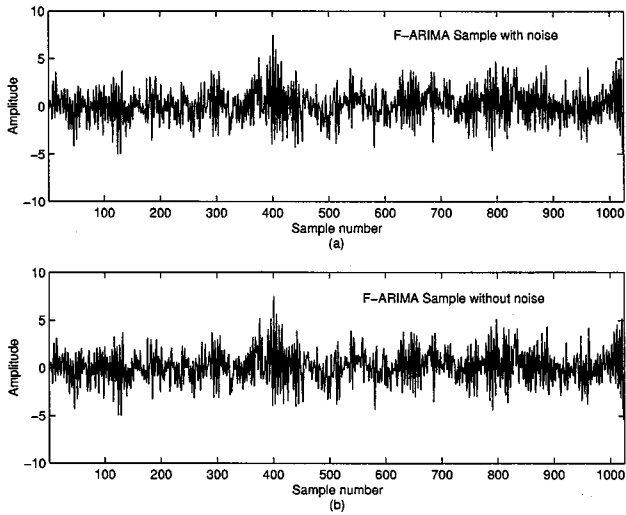


Fig. 6. Sample path of the F-ARIMA process in Example 2.

The standard deviation  $\sigma_w$  of the addition noise  $w(n)$  is set to be 0.2. A sample path of the F-ARIMA signal is plotted in Fig. 6. Let us denote the parameter vector  $\theta$  as

$$\begin{aligned} \theta &= [d \ \sigma_w \ a_1 \ a_2]^T \\ &= [\theta_1 \ \theta_2 \ \theta_3 \ \theta_4]^T \\ &= [0.35 \ 0.2 \ -0.4 \ 0.8]^T. \end{aligned}$$

The admissible set  $\Theta$  of  $\theta$  in this example is specified as

$$\Theta: \{\theta | 0 < \theta_1 < 0.5, \ 0 \leq \theta_2 \leq 1, \ -0.6 \leq \theta_3 \leq 0.6, \ -0.3 \leq \theta_4 \leq 1\}.$$

In order to meet the requirement of desired resolution for each parameter, all the bit numbers of the parameters  $\theta_i, i = 1, \dots, 4$  are assumed to be 16, i.e.,  $\ell_i = 16$ , for  $i = 1, \dots, 4$ . Therefore, the chromosome is of length  $\ell = \sum_{i=1}^4 \ell_i = 64$ , which leaves the search space  $\Theta$  with  $2^{64}$  alternatives.

To estimate the parameter vector  $\theta$  by employing the proposed GA-based method, the genetic parameters  $n_p, P_m, f_b,$  and  $f_w$  are specified as the same as those in Example 1. After 200 generations, the average estimated parameters are obtained as the following via Monte Carlo simulation with 100 runs:

$$\hat{d} = 0.3325, \ \hat{\sigma}_w = 0.1922, \ \hat{a}_1 = -0.4002, \ \hat{a}_2 = 0.8010.$$

The convergence of the mean log-likelihood function  $L_Y(\theta)$  is shown in Fig. 7, with quick convergence. The dash-dotted line in Fig. 7 indicates the theoretical maximal value of  $L_Y(\theta)$ . A typical trace, among 100 runs, of the parameter estimates is illustrated in Fig. 8.

The determination of the order of  $B(z)/A(z)$  can be obtained by solving the ML problem in (16) with different choices of  $n_a$  and  $n_b$ . The result is shown in Table II. We find that with the choice  $n_a = 2$  and  $n_b = 0$ , the likelihood function in (16) attains the maximum.

A comparison of the derived results with those obtained by using the ML-EM algorithm is presented in Table III via Monte Carlo simulation with 100 runs. It is shown that the proposed GA-based method exhibits a better performance than those of the ML-EM method with different initial conditions. Since there

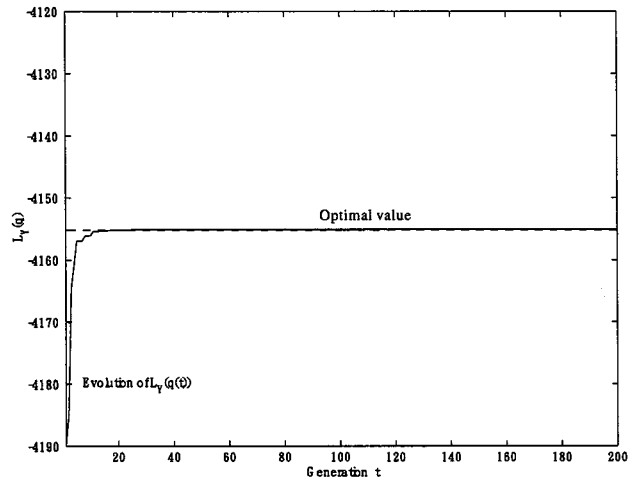


Fig. 7. Convergence of the average log-likelihood function versus generation in Example 2.

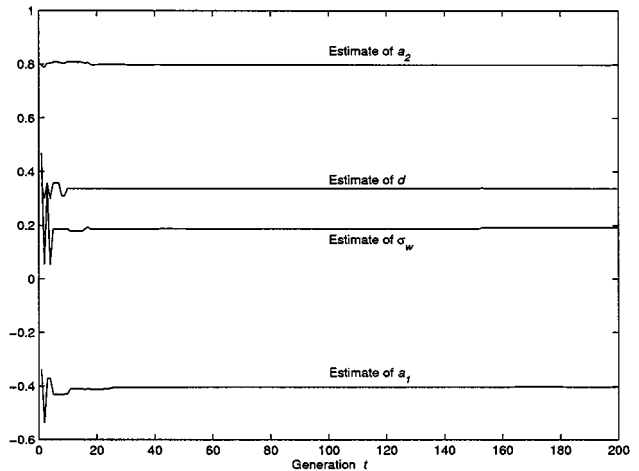


Fig. 8. Typical trace of parameter estimates in Example 2.

TABLE II  
MAXIMUM LIKELIHOOD  $L_Y(\theta)$  ATTAINED BY THE PROPOSED METHOD UNDER DIFFERENT MODEL ORDERS  $n_a$  AND  $n_b$  IN EXAMPLE 2

Model Order	$n_a = 0$ $n_b = 0$	$n_a = 1$ $n_b = 0$	$n_a = 1$ $n_b = 1$	$n_a = 2$ $n_b = 0$
Maximum likelihood max $L_Y(\theta)$	-4575.92	-4536.53	-4298.66	-4122.16
Model Order	$n_a = 2$ $n_b = 1$	$n_a = 2$ $n_b = 2$	$n_a = 3$ $n_b = 0$	$n_a = 3$ $n_b = 1$
Maximum likelihood max $L_Y(\theta)$	-4175.68	-4180.42	-4180.55	-4202.98

may be many local maxima near the optimal value, the ML-EM algorithm is usually trapped at a local maximum. Therefore, it is not easy to obtain a good estimate via the ML-EM method. However, it is seen that parameter estimation will be better if the initial value is chosen near the true value of the parameter.

*Example 3—A F-ARIMA Model:* The input fractal signal with length 1024 is the same as that in Example 1. The ARMA model is assumed to be

$$\frac{B(z)}{A(z)} = \frac{1 - z^{-1} + 0.85z^{-2}}{1 - 0.4z^{-1} + 0.8z^{-2}}.$$

TABLE III

MEAN ESTIMATES AND THEIR STANDARD DEVIATIONS (STDs) OF THE PARAMETER VECTOR  $[d \ \sigma_w \ a_1 \ a_2] = [0.35 \ 0.2 \ -0.4 \ 0.8]$  USING THE PROPOSED GA-BASED METHOD AND THE ML-EM METHOD VIA THE MONTE CARLO SIMULATION WITH 100 RUNS IN EXAMPLE 2

Parameter Estimation	Mean ( $\hat{d}$ , $\hat{\sigma}_w$ ) ( $\hat{a}_1$ , $\hat{a}_2$ )	STD (std $\hat{d}$ , std $\hat{\sigma}_w$ ) (std $\hat{a}_1$ , std $\hat{a}_2$ )
Estimation Method		
The proposed GA-based method	(0.3325, 0.1922) (-0.4002, 0.8010)	(0.1241, 0.0843) (0.1039, 0.0995)
ML-EM method with initial condition (0.34, 0.18, -0.38, 0.78)	(0.3401, 0.0881) (-0.3883, 0.7642)	(0.1353, 0.1746) (0.1473, 0.1183)
ML-EM method with initial condition (0.10, 0.01, -0.10, 0.10)	(0.2080, 0.0193) (-0.3463, 0.6131)	(0.2154, 0.2052) (0.1546, 0.1587)

Additive noise is assumed to be of standard deviation  $\sigma_w = 0.2$ . Let us denote  $\theta$  as

$$\begin{aligned} \theta &= [d \ \sigma_w \ a_1 \ a_2 \ b_1 \ b_2]^T \\ &= [\theta_1 \ \theta_2 \ \theta_3 \ \theta_4 \ \theta_5 \ \theta_6]^T \\ &= [0.35 \ 0.2 \ -0.4 \ 0.8 \ -1 \ 0.85]^T. \end{aligned}$$

The preassigned parameter space  $\Theta$  of  $\theta$  in this example is assumed to be

$$\Theta: \{\theta | 0 < \theta_1 < 0.5, \quad 0 \leq \theta_2 \leq 1 \\ -0.6 \leq \theta_3 \leq 0.6, \quad -0.3 \leq \theta_4 \leq 1 \\ -5 \leq \theta_5 \leq 5, \quad -5 \leq \theta_6 \leq 5\}.$$

All the required bit numbers of parameters are assumed to be 16 to meet the requirement of the desired resolution for each parameter. The search space is, therefore, with  $2^{96}$  alternatives. In this example, the genetic parameters  $P_m$ ,  $f_b$ , and  $f_w$  are the same as those in Example 1. The population size  $n_p$  is taken to be 400. After 200 generations via Monte Carlo simulation with 100 runs, the obtained average results are listed as

$$\begin{aligned} \hat{d} &= 0.3279, \quad \hat{\sigma}_w = 0.1668, \quad \hat{a}_1 = -0.3864 \\ \hat{a}_2 &= 0.7911, \quad \hat{b}_1 = -0.9355, \quad \hat{b}_2 = 0.7851. \end{aligned}$$

Convergence of the log-likelihood function with respect to the increasing of generation  $t$  is exhibited in Fig. 9, wherein the dash-dotted line shows the theoretical maximal value of the likelihood function. A typical trace, among 100 runs, of the parameter estimates is shown in Fig. 10. In comparison with the result obtained by using the ML-EM method, the derived results via Monte Carlo simulation with 100 runs are shown in Table IV. It is shown that the proposed GA-based method has a performance that is superior to the ML-EM method. The reason is that a local maximum result is obtained with the ML-EM method, but a near global optimal result is obtained with our proposed method.

*Remark 5:* From (14) and (15), it is seen that (14) is a more complicated function of the parameter vector than (15). Therefore, the manifold of the parameter space in the case without the ARMA part is less complex than that in the case with the AR or ARMA part. In this situation, it has a significant effect on parameter estimation. This effect can be found by viewing the simulation results in Figs. 4, 7, and 9, where the convergence rate in Fig. 4 is much faster than that in Figs. 7 and 9. The same

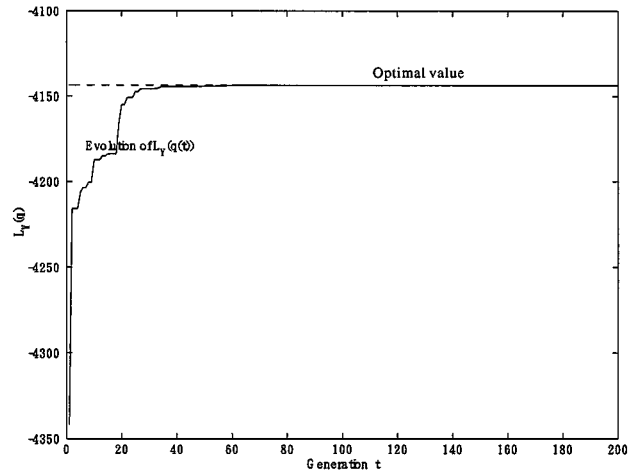


Fig. 9. Convergence of the average log-likelihood function versus generation in Example 3.

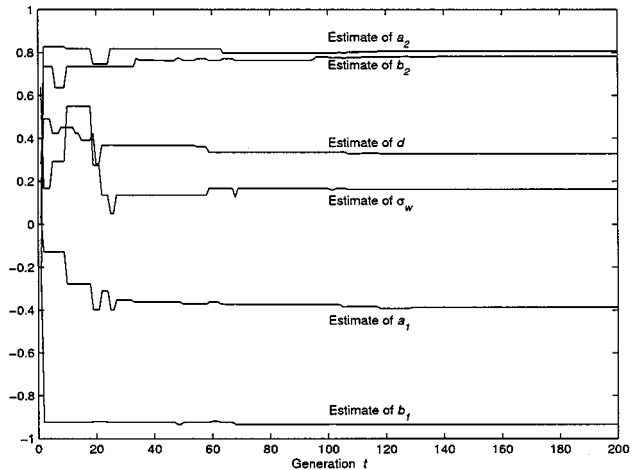


Fig. 10. Typical trace of parameter estimates in Example 3.

claim can also be obtained by observing the estimated results in Examples 1–3, where the derived result in Example 1 is more accurate than that in Examples 2 and 3.  $\square$

*Remark 6:* In the following, some brief comparisons of the proposed GA-based method with the ML-EM method are described.

The ML-EM method is an optimal estimation one for which the gradient-based optimization algorithm is used. In most of the gradient-based algorithms, the derivative information is usually

TABLE IV  
MEAN ESTIMATES AND THEIR STANDARD DEVIATIONS (STDs) OF THE PARAMETER VECTOR  $[d \ \sigma_w \ a_1 \ a_2 \ b_1 \ b_2] = [0.35 \ 0.2 \ -0.4 \ 0.8 \ -1 \ 0.85]$  USING THE PROPOSED GA-BASED METHOD AND THE ML-EM METHOD VIA THE MONTE-CARLO SIMULATION WITH 100 RUNS IN EXAMPLE 3

Parameter Estimation	Mean ( $\hat{d}, \hat{\sigma}_w, \hat{a}_1$ ) ( $\hat{a}_2, \hat{b}_1, \hat{b}_2$ )	STD (std $\hat{d}$ , std $\hat{\sigma}_w$ , std $\hat{a}_1$ ) (std $\hat{a}_2$ , std $\hat{b}_1$ , std $\hat{b}_2$ )
Estimation Method		
The proposed GA-based method	(0.3279, 0.1668, -0.3864) (0.7911, -0.9355, 0.7851)	(0.1015, 0.1020, 0.1034) (0.1058, 0.1105, 0.1005)
ML-EM method with initial condition (0.34, 0.18, -0.38, 0.78, -1.2, 0.95)	(0.3325, 0.0388, -0.3755) (0.7492, -1.1504, 0.9845)	(0.1327, 0.1565, 0.1418) (0.1296, 0.1330, 0.1241)
ML-EM method with initial condition (0.10, 0.01, -0.10, 0.10, -5, 2)	(0.1945, 0.0212, -0.3125) (0.6565, -2.1044, 1.1101)	(0.1523, 0.1664, 0.1631) (0.1600, 0.1539, 0.1470)

required. Many complicated computations, such as matrix inversion, are involved. By applying the ML-EM method to solve the optimal estimation problem addressed in this study from time domain perspective, the required computational time is very large, especially in the case of large amount of data points. However, a fractal signal is inherently a signal of long-range correlation structure. To capture this characteristic of a fractal signal, a large enough data of the observed signal is usually required. In this situation, ML-EM becomes inefficient. Moreover, ML-EM is an initial-condition dependent method owing to the use of the gradient-based algorithm. It is therefore easy to be trapped at local optima to obtain a suboptimal result, especially for highly nonlinear likelihood functions. Furthermore, the stability of the obtained result for the F-ARIMA model cannot be guaranteed by using the ML-EM method.

On the contrary, the proposed GA-based method is inherently a global optimization method. It is unnecessary to start with a good initial condition. Some simple operations such as string copying, string swapping, and bit changing are involved in the searching algorithm. Moreover, only some simple calculations are required in the computation of the log-likelihood function in the frequency domain. Because a stable search space is pre-specified, the stability of the obtained result for the F-ARIMA model is guaranteed. These characteristics make the proposed GA-based method in the frequency domain more suitable than the ML-EM method in the time domain to treat the parameter estimation problem for the F-ARIMA processes.  $\square$

## VI. CONCLUSIONS

A new class of the F-ARIMA processes is playing an increasingly important role in the area of signal processing. Accurate estimation of the parameters, especially the parameter  $d$ , is, however, important for practical applications. In this study, an ML estimation problem for estimating the parameters in the F-ARIMA processes has been proposed and efficiently solved in the frequency domain. The maximum of the highly nonlinear log-likelihood function is searched by employing the GA in the proposed estimation algorithm. Since the proposed algorithm simultaneously searches for many peaks and exchanges information among the peaks during the searching procedure, unlike other ML estimation methods, it possesses the property of global convergence in probability. Furthermore, stability of the F-ARIMA model is also guaranteed in

the parameter estimation procedure. The simulation results indicate that the proposed algorithm offers an effective and simple method to solve the nonlinear parameter estimation problem for the F-ARIMA processes. It has been shown that the proposed method is more initial-condition independent than the conventional ML-EM algorithm in solving the nonlinear parameter estimation problem. Therefore, the present results are believed to be useful for modeling and identification of the F-ARIMA processes before the design problems such as restoration, filtering, etc., are addressed.

## ACKNOWLEDGMENT

The authors are grateful to the anonymous referees, whose constructive and helpful comments led to significant improvements in the manuscript.

## REFERENCES

- [1] B. S. Chen and C. W. Lin, "Multiscale Wiener filter for the restoration of fractal signals: Wavelet filter bank approach," *IEEE Trans. Signal Processing*, vol. 42, pp. 2972–2982, Dec. 1994.
- [2] B. B. Mandelbrot, *The Fractal Geometry of Nature*. New York: Freeman, 1982.
- [3] M. F. Barnsley, *Fractals Everywhere*, 2nd ed. Boston, MA: Academic, 1993.
- [4] A. P. Pentland, "Fractal-based description of natural scenes," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-6, pp. 661–674, 1984.
- [5] M. Deriche and A. H. Tewfik, "Signal modeling with filtered discrete fractional noise processes," *IEEE Trans. Signal Processing*, vol. 41, pp. 2839–2849, Oct. 1993.
- [6] B. B. Mandelbrot and J. W. V. Ness, "Fractional Brownian motion, fractional noises, and application," *SIAM Rev.*, vol. 10, pp. 422–436, 1968.
- [7] G. W. Wornell and A. V. Oppenheim, "Estimation of fractal signals from noisy measurements using wavelets," *IEEE Trans. Signal Processing*, vol. 40, pp. 611–623, Mar. 1992.
- [8] J. H. Holland, *Adaptation in Natural and Artificial Systems*. Ann Arbor, MI: Univ. Michigan Press, 1975.
- [9] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*. Reading, MA: Addison-Wesley, 1989.
- [10] J. M. Fitzpatrick, J. J. Grefenstette, and D. V. Gucht, "Image registration by genetic search," in *Proc. IEEE Southeastcon.*, 1984, pp. 460–464.
- [11] D. M. Etter, M. J. Hicks, and K. H. Cho, "Recursive adaptive filter design using an adaptive genetic algorithm," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 1982, pp. 635–638.
- [12] K. Kristinsson and G. A. Dumont, "Genetic algorithms in system identification," in *Proc. Third IEEE Int. Symp. Intell. Contr.*, Arlington, VA, 1988, pp. 597–602.
- [13] C. L. Karr and E. J. Gentry, "Fuzzy control of pH using genetic algorithms," *IEEE Trans. Fuzzy Syst.*, vol. 1, pp. 46–53, Feb. 1993.
- [14] G. Rudolph, "Convergence analysis of canonical genetic algorithms," *IEEE Trans. Neural Networks*, vol. 5, pp. 96–101, Jan. 1994.

- [15] L. Yao and W. A. Sethares, "Nonlinear parameter estimation via the genetic algorithm," *IEEE Trans. Signal Processing*, vol. 42, pp. 927–935, Apr. 1994.
- [16] J. M. Mendel, *Lessons in Digital Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1987.
- [17] J. H. Holland, "Outline for a logical theory of adaptive systems," *J. ACM*, vol. 3, pp. 297–314, 1962.
- [18] M. F. Barnsley, R. L. Devaney, B. B. Mandelbrot, H. O. Peitgen, D. Saupe, and R. F. Voss, *The Science of Fractal Images*. New York: Springer-Verlag, 1988.
- [19] A. J. Crilly, R. A. Earnshaw, and H. Jones, Eds., *Fractals and Chaos*. New York: Springer Verlag, 1991.
- [20] J. Beran, *Statistics for Long-Memory Processes*. London, U.K.: Chapman & Hall, 1994.
- [21] S. M. Kay, *Modern Spectral Estimation: Theory and Application*. Englewood Cliffs, NJ: Prentice-Hall, 1988.
- [22] D. R. Brillinger, *Time Series Data Analysis and Theory*. New York: Holt, Rinehart, and Winston, Inc., 1975.
- [23] C. W. Granger and R. Joyeux, "An introduction to lone memory time series models and fractional differencing," *J. Time Series Anal.*, vol. 1, 1980.
- [24] J. R. M. Hosking, "Fractionally differencing," *Biometrika*, vol. 68, pp. 165–176, 1981.
- [25] W. E. Leland, M. S. Taggu, W. Willinger, and D. V. Wilson, "On the self-similar nature of Ethernet traffic," *IEEE/ACM Trans. Networking*, vol. 2, pp. 1–15, 1994.
- [26] A. Adas, "Traffic models in broadband networks," *IEEE Commun. Mag.*, vol. 35, pp. 82–89, July 1997.
- [27] J. Beran *et al.*, "Long-range dependence invariable-bit-rate video traffic," *IEEE Trans. Commun.*, vol. 43, pp. 1566–1579, July 1995.
- [28] G. N. Fouskitakis and S. D. Fassios, "Pseudolinear estimation of fractionally integrated ARMA (ARFIMA) models with automotive application," *IEEE Trans. Signal Processing*, vol. 47, pp. 3365–3380, Dec. 1999.
- [29] V. J. Ribeiro, R. H. Riedi, M. S. Crouse, and R. G. Baraniuk, "Multiscale queuing analysis of long-range-dependent network traffic," in *Proc. Nineteenth Annu. Joint Conf. IEEE Comput. Commun. Soc.*, vol. 2, 2000, pp. 1026–1035.
- [30] P. Goncalves, R. H. Riedi, and R. Baraniuk, "A simple statistical analysis of wavelet-based multifractal spectrum estimation," in *Proc. Thirty-Second Asilomar Conf. Signals, Syst., Comput.*, vol. 1, 1998, pp. 287–291.
- [31] R. H. Riedi, M. S. Crouse, V. J. Ribeiro, and R. G. Baraniuk, "A multifractal wavelet model with application to network traffic," *IEEE Trans. Inform. Theory*, vol. 45, pp. 992–1018, June 1999.
- [32] R. M. Dufour and E. L. Miller, "Image restoration with  $1/f$  type fractal models and statistical estimation of model parameters," in *Proc. Int. Conf. Image Process.*, vol. 2, 1998, pp. 88–92.
- [33] —, "Statistical estimation of  $1/f$  type prior models: Robustness to mismatch and efficient model determination," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 5, 1996, pp. 2491–2494.
- [34] K. J. Astrom and B. Wittenmark, *Computer Controlled Systems: Theory and Application*. Englewood Cliffs, NJ: Prentice-Hall, 1984.
- [35] V. K. Rohatgi, *An Introduction to Probability Theory and Mathematical Statistics*. New York: Wiley, 1975.
- [36] Y. Yajima, "A central limit theorem of Fourier transforms of strongly dependent stationary processes," *J. Time Series Anal.*, vol. 10, no. 4, pp. 375–383, 1989.



**Bor-Sen Chen** (M'82–SM'89–F'01) received the B.S. degree from Tatung Institute of Technology, Taipei, Taiwan, R.O.C., the M.S. degree from National Central University, Chungli, Taiwan, and the Ph.D. degree from the University of Southern California, Los Angeles, in 1970, 1973, and 1982, respectively.

From 1973 to 1987, he was a Lecturer, Associate Professor, and Professor at Tatung Institute of Technology. Currently, he is a Professor with National Tsing Hua University, Hsinchu, Taiwan. His current research interests include control signal processing and bioinformatics. He is an Editor of the *Journal of Asian Control*.

Dr. Chen has received the Distinguished Research Award from National Science Council of Taiwan four times. He is a Research Fellow of the National Science Council and the Chair of the Outstanding Scholarship Foundation. He was the Chairman of Taipei Chapter of IEEE Control Society in 1991. He is now an Associate Editor of IEEE TRANSACTIONS ON FUZZY SYSTEMS.



**Bore-Kuen Lee** (S'87–M'91) was born in Taichung, Taiwan, R.O.C., in 1961. He received the M.S. and Ph.D. degrees in automatic control engineering from National Chiao Tung University, Hsinchu, Taiwan, in 1986 and 1991, respectively.

From 1986 to 1987, he worked in the robotic control area as an Associate Researcher at the Mechanical Industry Research Laboratories, the Industrial Technology Research Institute, Hsinchu, Taiwan. Since 1991, he has been with the Department of Electrical Engineering, Chung Hua University, Hsinchu, Taiwan, where he is currently an Associate Professor. His research interests include variable structure control theory, estimation theory, fuzzy control, and traffic control of networks.

Dr. Lee is a member of the Phi Tau Phi society.



**Sen-Chueh Peng** (M'92) was born in Kaohsiung, Taiwan, R.O.C. He received the B.S. degree from Kaohsiung Institute of Technology and the M.S. and Ph.D. degrees in electrical engineering from National Tsing-Hua University, Hsinchu, Taiwan, in 1986 and 1992, respectively.

From 1992 to 1998, he was an Associate Professor at National Yun-Lin Polytechnic Institute, Yunlin, Taiwan. He is now a Professor at National Huwei Institute of Technology, Yunlin, Taiwan. His areas of research interest are signal processing,

communication, and robust control.