Multi-modulus algorithm based on global artificial fish swarm intelligent optimization of DNA encoding sequences

Y.C. Guo[1,2], H. Wang[1], H.P. Wu[1] and M.Q. Zhang[1]

[1]Jiangsu Key Laboratory of Meteorological Observation and Information Processing, Nanjing University of Information Science & Technology, Nanjing, Jiangsu, China
[2]Jiangsu Collaborative Innovation Center on Atmospheric Environment and Equipment Technology, Nanjing, China

Corresponding author: Y.C. Guo
E-mail: guo-yecai@163.com

ABSTRACT. Aimed to address the defects of the large mean square error (MSE), and the slow convergence speed in equalizing the multi-modulus signals of the constant modulus algorithm (CMA), a multi-modulus algorithm (MMA) based on global artificial fish swarm (GAFS) intelligent optimization of DNA encoding sequences (GAFS-DNA-MMA) was proposed. To improve the convergence rate and reduce the MSE, this proposed algorithm adopted an encoding method based on DNA nucleotide chains to provide a possible solution to the problem. Furthermore, the GAFS algorithm, with its fast convergence and global search ability, was used to find the best sequence. The real and imaginary parts of the initial optimal weight vector of MMA were obtained through DNA coding of the best sequence. The simulation results show that the proposed algorithm has a faster convergence

speed and smaller MSE in comparison with the CMA, the MMA, and the AFS-DNA-MMA.

Key words: DNA encoding; Global artificial fish swarm algorithm; Multi-modulus algorithm

INTRODUCTION

To eliminate inter-symbol interference (ISI) and improve the performance of the communication system, some blind equalization algorithms are used in the receiving terminal to compensate and eliminate the ISI of the channels. In the known blind equalization algorithms, traditional constant modulus algorithm (CMA) has lower complexity and stronger robustness, yet it has low convergence speed, local convergence, and large mean square error (MSE). Furthermore, it does not effectively equalize the higher-order multi-modulus quadrature amplitude modulation (QAM) signals (Yuan et al., 2012). The QAM signals have a plural modulus value and the signal constellations are distributed in circles with different radii. Therefore, the multi-modulus algorithm (MMA) can effectively equalize higher-order multi-modulus QAM signals (Paracha and Zerguine, 2011).

The focus of DNA coding is to avoid hybridization of biochemical reactions among DNA molecules as much as possible. Currently, many algorithms can optimize DNA encoding sequences, and intelligent algorithms such as the genetic algorithm (GA) (Vijayaraghavan et al., 2013; Garg et al., 2014a,b,c) and the artificial fish swarm algorithm (AFSA) have been studied (Li and Suohai, 2013). The global AFSA (GAFSA) is an efficient intelligence algorithm that has global search ability and can overcome the low accuracy, slow convergence, and high complexity of the AFSA.

In this paper, based on the advantages of DNA coding, MMA, and GAFSA, the MMA, which is based on the global artificial fish swarm (GAFS) intelligent optimization of DNA encoding sequences (GAFS-DNA-MMA), is proposed to overcome the defects of the CMA and the MMA.

MATERIAL AND METHODS

Multi-modulus algorithm

Output signals of the CMA are equalized to a circle with the radius R, which is a statistical modulus value of the transmitted signals. The constellation points of the higher-order multi-modulus QAM signals are distributed on several circles of different radii. When we use the CMA to equalize higher-order QAM signals, the output constellations of the CMA are equalized to a fixed circle, so there is a larger MSE. The MMA, based on amplitude and phase information, replaces modulus values of in-phase and quadrature components with real and imaginary parts of the input signals, respectively. The principle is shown in Figure 1, and can effectively correct phase rotation; however, the MMA has some defects such as slow convergence rate and large MSE (Ni et al., 2012).
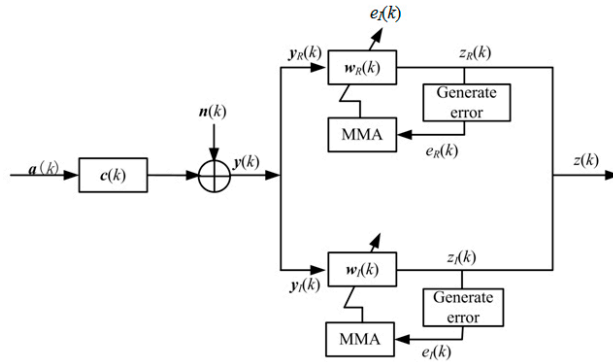
Figure 1. Principle of the multi-modulus algorithm (MMA).

In Figure 1, $a(k)$ is a zero mean independent identically distributed transmitted signal vector, $c(k)$ denotes the channel impulse response vector, and $n(k)$ is an additive white Gaussian noise vector, $y(k)$ represents the signals by the equalizer, $y_R(k)$ and $y_I(k)$ are the real and imaginary parts of $y(k)$, $f(k)$ is the equalizer weight vector, $f_R(k)$ and $f_I(k)$ are the real and imaginary parts of $f(k)$. $z(k)$ is the equalizer output signal, $z_R(k)$ and $z_I(k)$ are the real and imaginary parts of $z(k)$. $e(k)$ is the error term, $e_R(k)$ and $e_I(k)$ are the real and imaginary parts of $e(k)$.

The cost function of the MMA is given by

$$J_{MMA} = E[e_R^{\,2}(k)] + E[e_I^{\,2}(k)] \qquad \text{(Equation 1)}$$

The corresponding iterative formula of tap coefficients in the MMA is written as

$$\begin{cases} f_R(k+1) = f_R(k) - \mu e_R(k)\mathbf{y}_R(k) \\ f_I(k+1) = f_I(k) - \mu e_I(k)\mathbf{y}_I(k) \end{cases} \qquad \text{(Equation 2)}$$

where $e_R(n) = z_R(n)(z_R^{\,2}(n) - R_R)$, $e_I(n) = z_I(n)(z_I^{\,2}(n) - R_I)$, μ is an iterative step-size.

Global artificial fish swarm algorithm

The GAFSA is an efficient intelligence algorithm, which can imitate the behavior of fish swarms, including preying, swarming, and rear-end collisions, to achieve the global optimal value (Ma and Wang, 2009; Gao and Chen, 2010).

Preying behavior: Let $\mathbf{v}_i$ be the current position vector of the global artificial fish and a position vector $\mathbf{v}_j$ in current visual distances is selected randomly, we use $Y_1(\mathbf{v}_i)$ to denote the fitness function of the global artificial fish. If the fitness function is value $Y_1(\mathbf{v}_j) > Y_1(\mathbf{v}_i)$, then the global artificial fish moves a step from position $\mathbf{v}_i$ to position $\mathbf{v}_j$ according to Equation 3, otherwise, the global artificial fish selects another position vector $\mathbf{v}_j$ and judges whether it can reach the forward movement condition. If it cannot reach the forward movement condition, it randomly moves a step forward according to Equation 4.

$$v_{i+1} = v_i + \frac{(v_j - v_i) + (v_{best} - v_i)}{\left\| (v_j - v_i) + (v_{best} - v_i) \right\|} \cdot step \cdot \text{rand}(0,1) \qquad \text{(Equation 3)}$$

$$v_{i+1} = v_i + V \cdot \text{rand}(0,1) \qquad \text{(Equation 4)}$$

where $\text{rand}(0,1)$ means random number within 0 to 1, *step* is a step-size, V is a visual range, v_{best} is the best position vector.

Swarming behavior: Assume that the current position vector v_i of the global artificial fish, the center position vector v_c in the current visual range, and the number of the companions of the artificial fishes are searched. If the fitness function value is $Y_1(v_c) > Y_1(v_i)$, the center position has more food and artificial fishes are not too crowded. Then the artificial fish goes a step towards the center position according to Equation 5. Otherwise, the artificial fish continues performing preying behavior according to Equations 3 or 4.

$$v_{i+1} = v_i + \frac{(v_c - v_i) + (v_{best} - v_i)}{\left\| (v_c - v_i) + (v_{best} - v_i) \right\|} \cdot step \cdot \text{rand}(0,1) \qquad \text{(Equation 5)}$$

Following behavior: Assume that v_{max} denotes the position vector corresponding to the maximum fitness function value in the current visual distance. If the number of companions of the artificial fishes in the visual distance meets the condition for $N_v / N > \delta$, where δ denotes the degree of congestion and N_v is the number of companions within the visual distance. If the fitness function value is $Y_1(v_{max}) > Y_1(v_i)$, the artificial fish goes a step towards position vector v_{max} according to Equation 6. Otherwise, the artificial fish performs swarming behavior as shown in Equation 5.

$$v_{i+1} = v_i + \frac{(v_{max} - v_i) + (v_{best} - v_i)}{\left\| (v_{max} - v_i) + (v_{best} - v_i) \right\|} \cdot step \cdot \text{rand}(0,1) \qquad \text{(Equation 6)}$$

MMA based on optimization of GAFS DNA encoding sequence

Hamming distance constraint based on DNA encoding

The hamming distance between two DNA sequences is the sum of the different characters in all corresponding positions that describe that the two sequences are not similar (Cui et al., 2007; Wang and Geng, 2012). DNA sequence X and Z are denoted as $X = x_1 x_2 \cdots x_n$ and $Z = z_1 z_2 \cdots z_n$, where n is the number of DNA bases in X or Z, x_m and z_l are any of the four bases (adenine, guanine, cytosine, and thymine). Their hamming distance is written as $H(X,Z)$ and defined as

$$H(X,Z) = \sum_{i=1}^{n} h(x_i, z_i) \qquad \text{(Equation 7)}$$

where $h(x_i,z_i) = \begin{cases} 0, if\ x_i = z_i \\ 1, if\ x_i \neq z_i \end{cases}$.

In the DNA encoding design, the number of different bases in sequence X and Y increases with an increase in hamming distance. Therefore, the possibility of hybridization is reduced (Tulpan et al., 2014).

Fitness functions

In this section, we define two fitness functions.

The first fitness function handles the question using the weighted average method according to Equation 8:

$$Y_1(s_i) = wF(s_i,s_i) = w \cdot \min_{-n<k<n} H(s_i, \sigma^k(s_i^R)) \qquad \text{(Equation 8)}$$

where s_i is the i th DNA sequence and equals to the i th position vector v_i, if $k>0(<0)$, $\sigma^k(s_i^R)$ denotes the sequence s_i^R shift to right (left). s_i^R is the anti-chain of s_i. k is the number of shifts and w is set to 1 to enable convenient calculation.

The second fitness function of the AFSA is defined by the cost function of the MMA and is written as

$$Y_2(v_i) = \frac{1}{J_{MMA}(f_i)} = \frac{1}{E[e_R^2(k)] + E[e_I^2(k)]} \qquad \text{(Equation 9)}$$

where $J_{MMA}(f_i)$ denotes the cost function of MMA.

RESULTS

Steps of the proposed algorithm

Step 1: Initialization of DNA sequence. The initial population of DNA sequence $s = [s_1, s_2, \cdots, s_N]$, where s_m is the m th DNA sequence and N is the number of DNA sequences. It is equal to the initial position vector of GAFS.

Step 2: The calculation of first fitness function. The first fitness function of each artificial fish is calculated according to Equation 8. The maximum value of the first fitness function and its corresponding position vector are recorded in the first bulletin board.

Step 3: DNA coding and calculation of the second fitness function. The position vector with DNA coding of the GAFS is obtained. The value of the second fitness function for each artificial fish swarm is calculated according to Equation 9, and the maximum value and its corresponding position vector are recorded in the second bulletin board.

Step 4: The preying, swarming, or following behavior of artificial fish. Each artificial fish in the swarm carries out following behavior according to Equation 6. If the following behavior does not occur, all fishes will carry out swarming behavior according to Equation 5. If swarming behavior is not successful, all fishes will carry out preying behavior according to Equations 3 and 4. Each behavior can cause the current position vector to change.

Step 5: Update of the first bulletin board. The first fitness function of each artificial fish is calculated. The maximum value is selected and compared with the previous maximum value of first fitness function recorded in the bulletin board. If the maximum value is greater than the previous maximum value, the first bulletin board is updated by the current maximum value and its corresponding position vector.

Step 6: Coding DNA and the update of the second bulletin board. After one iteration, the position vector in the first bulletin board is obtained by the coding DNA. The second fitness function of each artificial fish is calculated. The maximum value is selected and compared with the previous maximum value of the second fitness function recorded in the second bulletin board. If the maximum value is greater than the previous maximum value recorded in the second bulletin board, the second bulletin board is updated by the current maximum value and its corresponding position vector.

Step 7: Determination of the termination condition. Judge whether the iteration is equal to the maximum value. If the condition does not hold, the iteration process returns to step 4; otherwise, the record in the second bulletin board is the output.

Step 8: The record in the second bulletin board is used as the real and imaginary parts of the initial optimization weight vector and is updated according to Equation 2.

Simulation analyses

In the simulations, the transmitted signals were 16 QAM signals, the channel was $h = [0.9656\ -0.0906\ 0.0578\ 0.2368]$, signal-to-noise ratio was 20 dB, the number of equalizer weight coefficients was $L = 11$, and the step-sizes of the AFS-DNA-MMA and the GAFS-DNA-MMA was $\mu = 0.002$. The step-size of the MMA was $\mu = 0.01$. The number of GAFSA was $N = 110$, the crowding factor was $\delta = 0.5$. The simulation results are shown in Figure 2.
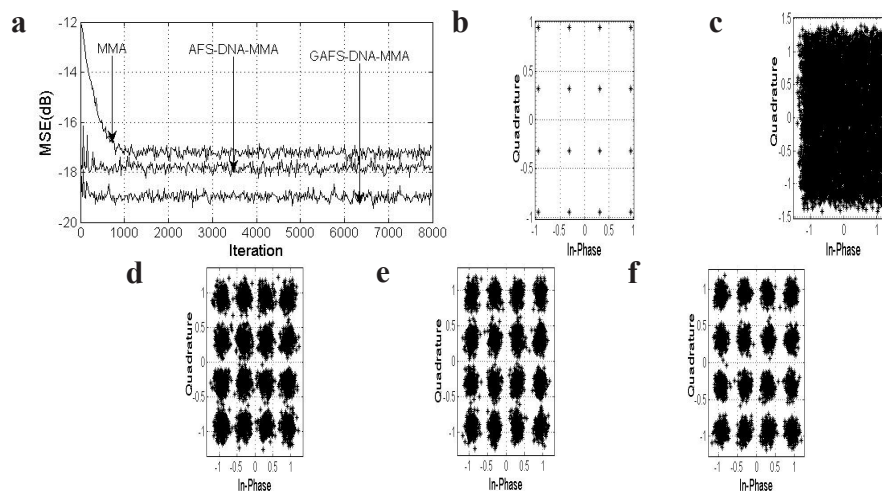

Figure 2. Simulated results. **a.** The curves of MSE. **b.** The constellations of the transmitted signals. **c.** The input constellations of the equalizer. **d.** The output constellations of the MMA. **e.** The output constellations of the AFS-DNA-MMA. **f.** The output constellations of the GAFS-DNA-MMA. MSE, mean square error; MMA, multi-modulus algorithm; AFS-DNA-MMA, multi-modulus algorithm based on artificial fish swarm intelligent optimization of DNA encoding sequences; GAFS-DNA-MMA, multi-modulus algorithm based on global artificial fish swarm intelligent optimization of DNA encoding sequences.

From Figure 2, we can see that the GAFS-DNA-MMA outperforms the AFS-DNA-MMA and MMA in improving the convergence speed and reducing MSE. The MSE of the GAFS-DNA-MMA is smallest and its convergence speed is fastest. Moreover, the output constellations of the GAFS-DNA-MMA are more compact and clearer than the MMA and have no phase rotation.

DICUSSION

After DNA encoding and the GAFSA are applied to the MMA to optimize its initial weight vector, the GAFS-DNA-MMA is proposed. The results of the simulation show that the MSE of the GAFS-DNA-MMA has the fastest convergence rate and the lowest MSE, as well as more compact and clearer output constellations in comparison with the AFS-DNA-MMA and MMA. Compared with the global artificial fish swarm algorithm based on the DNA genetic algorithm (GA-GAFSA), the GAFSA takes longer to find the optimal value. Therefore, in future, we can investigate the GA-GAFSA in combination with the MMA. Alternatively, the current novel methodology can also be compared with soft computing methods such as genetic programming.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

REFERENCES

Cui G, Cao X, Zhou J and Wang Y (2007). The optimization of DNA encoding sequences based on improved AFS algorithms. IEEE International Conference on Automation and Logistics, 1141-1144.

Gao Y and Chen Y (2010). The optimization of water utilization based on artificial fish-swarm algorithm. *Fifth Int. Conf. Nat. Comp.* 8: 4415-4419.

Garg A, Garg A, Tai K, Barontini S, et al. (2014a). A computational intelligence-based genetic programming approach for the simulation of soil water retention curves. *Transport Porous Med.* 103: 497-513.

Garg A, Tai K and Gupta AK (2014b). A modified multi-gene genetic programming approach for modelling true stress of dynamic strain aging regime of austenitic stainless steel 304. *Meccanica* 49: 1193-1209.

Garg A, Tai K and Savalani MM (2014c). Formulation of bead width model of an SLM prototype using modified multi-gene genetic programming approach. *Int. J. Adv. Manuf. Tech.* 73: 375-388.

Li M and Suohai F (2013). Forex prediction based on SVR optimized by artificial fish swarm algorithm. Fourth Global Congress on Intelligent Systems, 47-52.

Ma H and Wang Y (2009). An artificial fish swarm algorithm based on chaos Search. *Fifth Int. Conf. Nat. Comp.* 4: 118-121.

Ni Y, Du D, Xiao R and Xie C (2012). Multi-modulus blind equalization algorithm based on high-order QAM genetic optimization. Eighth International Conference on Natural Computation, 679-682.

Paracha KN and Zerguine A (2011). A Newton-like algorithm for adaptive multi-modulus blind equalization. 7th International Workshop on Systems, Signal Processing and their Applications, 283-286.

Tulpan D, Smith DH and Montemanni R (2014). Thermodynamic post-processing versus GC-content preprocessing for DNA codes satisfying the hamming distance and reverse-complement constraints. *IEEE/ACM T Comput. Bi.* 11: 441-452.

Vijayaraghavan V, Garg A, Wong CH and Tai K (2013). Estimation of mechanical properties of nanomaterials using artificial intelligence methods. *Appl. Phys. A-Mater.* 116: 1099-1107.

Wang Q and Geng X (2012). Application research of biosensor technology in DNA computing. *IEEE Int. Conf. Comp. Sci. Automation Eng.* 1: 639-643.

Yuan JT, Chao JH and Lin TC (2012). Effect of channel noise on blind equalization and carrier phase recovery of CMA and MMA. *IEEE Trans. Commun.* 60: 3274-3285.