Fractional-order Backpropagation Neural Networks: Modified Fractional-order Steepest Descent Method for Family of Backpropagation Neural Networks

Yi-Fei PU, and Jian WANG.

Abstract—This paper offers a novel mathematical approach, the modified Fractional-order Steepest Descent Method (FSDM) for training Backpropagation Neural Networks (BPNNs); this differs from the majority of the previous approaches and as such. A promising mathematical method, fractional calculus, has the potential to assume a prominent role in the applications of neural networks and cybernetics because of its inherent strengths such as long-term memory, nonlocality, and weak singularity. Therefore, to improve the optimization performance of classic first-order BPNNs, in this paper we study whether it could be possible to modified FSDM and generalize classic first-order BPNNs to modified FSDM based Fractional-order Backpropagation Neural Networks (FBPNNs). Motivated by this inspiration, this paper proposes a state-of-the-art application of fractional calculus to implement a modified FSDM based FBPNN whose reverse incremental search is in the negative directions of the approximate fractional-order partial derivatives of the square error. At first, the theoretical concept of a modified FSDM based FBPNN is described mathematically. Then, the mathematical proof of the fractional-order global optimal convergence, an assumption of the FSDM based FBPNN whose reverse incremental search is in the negative state-of-the-art application of fractional calculus to implement a modified FSDM based FBPNN is described mathematically. Finally, we perform comparative experiments and compare a modified FSDM based FBPNN with a classic first-order BPNN, i.e., an example function approximation, fractional-order multi-scale global optimization, and two comparative performances with real data. The more efficient optimal searching capability of the fractional-order multi-scale global optimization of a modified FSDM based FBPNN is analysed in detail. Finally, we perform comparative experiments and compare a modified FSDM based FBPNN with a classic first-order BPNN, i.e., an example function approximation, fractional-order multi-scale global optimization, and two comparative performances with real data. The more efficient optimal searching capability of the fractional-order multi-scale global optimization of a modified FSDM based FBPNN is determined by the global optimal solution is the major advantage being superior to a classic first-order BPNN.

Index Terms—Fractional calculus, Fractional-order backpropagation algorithm, Fractional-order steepest descent method, Mean squared error, Fractional-order multi-scale global optimization searching

I. INTRODUCTION

BACKPROPAGATION, an abbreviation for “backward propagation of errors”, is a common method for training artificial neural networks and is used in conjunction with an optimization method such as gradient descent [1]. The classic first-order Backpropagation Algorithm (BA) to train multilayer networks was first described in the thesis of P. J. Werbos in 1974; the algorithm was presented in the context of general networks, with neural networks as a special case [2]. It was not until the mid-1980s that the classic BA was widely publicized. It was rediscovered independently by D. E. Rumelhart et al. [3], D. B. Parker [4], and Y. L. Cun [5]. The classic BA was popularized by its inclusion in the book Parallel Distributed Processing [6]. The publication of this book triggered a significant amount of research into neural networks. Multilayer perceptrons, trained by the classic BA, are currently the most widely used neural networks. The classic BA can refer to the result of a playout that is propagated up the search tree in a Monte Carlo tree search [7]. It has been demonstrated that classic two-layer first-order Backpropagation Neural Networks (BPNNs), with sigmoid activation functions in the hidden layer and linear transfer functions in the output layer, can approximate any function of interest to any degree of accuracy when a sufficient number of hidden units are available [8]. Research related to faster BPNN algorithms can be broadly classified into two categories. The first category involves the development of heuristic techniques. These heuristic techniques include concepts such as varying the learning rate using momentum and resealing variables [9]–[12]. The second category focuses on standard numerical optimization techniques. These techniques include concepts such as the conjugate gradient algorithm and Levenberg-Marquardt algorithm [13]–[17]. Combinations of neural networks and evolutionary computation procedures have been widely explored. This area addresses a wide range of topics such as cutting angle method [18], simulated annealing [19]–[21], swarm algorithms [22], genetic algorithms [23]–[26], and hybrid training methods [27]–[28], which are superior to traditional local techniques. In all these methods, the computational complexity increases rapidly with an increase in the number of variables. Moreover, the classic first-order BA-based BPNNs are easily trapped into a local optimal solution, whose optimization performance must be improved.

The application of fractional calculus to neural networks and cybernetics is an emerging discipline of research and a small number of studies have been conducted in this area. Fractional calculus has become an important novel branch in mathematical analyses [29], [30]. Recently, fractional calculus has become a promising mathematical method for physical scientists and engineering technicians. Promising results and ideas have demonstrated that fractional calculus can be an interesting and useful tool in many scientific fields such as diffusion processes [31], viscoelasticity theory [32], fractal dynamics [33], fractional control [34], [35], fractance [36], fracmennister [37], image processing [38], [39], and neural networks [40]–[43]. Fractional calculus has been applied to neural networks and cybernetics primarily owing to its inherent advantages of long-term memory, nonlocality, and weak singularity, important properties of fractional calculus [29], [30]. The basic characteristic feature of fractional calculus is that it extends the concepts of the integer-order difference and Riemann sums. The characteristics of fractional calculus are considerably different from those of classic integer-order calculus. For example the fractional differential, except based on the Caputo definition, of a Heaviside function is nonzero, whereas its integer-order differential must be zero [29], [30]. Thus, the properties of the Fractional-order Steepest Descent Method (FSDM) are also different from those of the traditional first-order steepest descent method [40]. For example, the FSDM can determine the fractional-order extreme points of the energy norm, which do not overlap the traditional first-order stationary points [40]. It is known that classic first-order BPNNs demonstrate a tendency to be trapped into a local
optimal solution. The application of the modified FSDM to training BPNNs has the potential of overcoming such deficits. Therefore, to improve the optimal searching capability of traditional first-order BPNNs, it is logical to modify FSDM and determine a method to generalize a classic first-order BPNN to a modified FSDM based FBPN. Based on this inspiration, in this study, we introduce a modified FSDM based FBPN, whose reverse incremental search is in the negative directions of the approximate fractional-order partial derivatives of the square error at iteration $k$ of the iterative search process. This paper offers a novel mathematical approach, the modified FSDM, for training BPNNs, which differs from the majority of the previous approaches and as such, represents an interesting theoretical contribution. The more efficient optimal searching capability of the fractional-order multi-scale global optimization of a modified FSDM based FBPN to determine the global optimal solution is the major advantage being superior to a classic first-order BPNN.

The remainder of this manuscript is organized as follows: In Section 2, the necessary theoretical background for fractional calculus and the FSDM is presented. In Section 3, the theoretical concept of a modified FSDM based FBPN is discussed in detail. At first, we describe the fractional-order partial derivatives of the square error of a BPNN. Secondly, the achievement of the modified FSDM for the family of the BPNNs and the modified FSDM based FBPNs is further discussed. Thirdly, the mathematical proof of the fractional-order global optimal convergence, an assumption of the structure, and the fractional-order multi-scale global optimization of a modified FSDM based FBPN are analysed. The experimental results and analysis are reported in Section 4. Section 4 presents a sample function approximation, a fractional-order multi-scale global optimization, and two comparative performances with real data. In Section 5, the conclusions of this manuscript are presented.

II. RELATED WORK

This section includes a brief introduction to the necessary mathematical background of fractional calculus and the FSDM.

The commonly used fractional calculus definitions are Grünwald-Letnikov, Riemann-Liouville, and Caputo [29], [30]. The Grünwald-Letnikov definition of fractional calculus for causal signal $f(x)$ can be represented in a convenient form as follows:

$$e^{-x}D_x^v f(x) = \lim_{h \to 0} \left( (x - a) / h \right)^{-\xi} \left( (k - v) / h \right)^{\xi} \left( x - a \right)^{-\xi} \left( (x - a) / h \right)^{\xi} \left( x - a \right)^{-\xi}$$

where $f(x)$ is a differentiable function [29], [30], $[a, x]$ is the duration of $f(x)$, $N$ is the number of partitions of the duration, $v$ is an arbitrary real number, $\Gamma(x) = \int_a^x e^{-x}x^\xi dx$ is gamma function and $e^{-x}D_x^v$ denotes the Grünwald-Letnikov defined fractional differential operator. In this manuscript, we use the equivalent notations $D_x^v = e^{-x}D_x^v$ in an interchangeable manner.

Secondly, the reverse incremental search of the FSDM is in the negative direction of the $\nu$-order fractional derivative of the quadratic energy norm $E$, which can be represented as [40]:

$$\xi_{i+1} = \xi_i - \mu (D_{x}^{\nu} E)$$

where $\mu$ is the step size or number of iterations, $\xi$ is the independent variable of $E$, $D_{x}^{\nu} E$ is the $\nu$-order fractional derivative of $E$ at $\xi = \xi_i$, and $\mu$ is the constant coefficient that controls the stability and the rate of convergence of the FSDM.

III. FRACTIONAL-ORDER BACKPROPAGATION NEURAL NETWORKS

A. Fractional-order partial derivatives of square error

In this subsection, to achieve the modified FSDM for the family of the BPNNs, the fractional-order partial derivatives of the square error of a BPNN should be described firstly.

Figure 1 displays the model of a BPNN, which is represented by abbreviated symbols denoting its three layers. The remainder of this manuscript is organized as follows: In Section 2, the necessary theoretical background for fractional calculus and the FSDM is presented. In Section 3, the theoretical concept of a modified FSDM based FBPN is discussed in detail. At first, we describe the fractional-order partial derivatives of the square error of a BPNN. Secondly, the achievement of the modified FSDM for the family of the BPNNs and the modified FSDM based FBPNs is further discussed. Thirdly, the mathematical proof of the fractional-order global optimal convergence, an assumption of the structure, and the fractional-order multi-scale global optimization of a modified FSDM based FBPN are analysed. The experimental results and analysis are reported in Section 4. Section 4 presents a sample function approximation, a fractional-order multi-scale global optimization, and two comparative performances with real data. In Section 5, the conclusions of this manuscript are presented.
In addition, \( \sum_{i=0}^{n} \sum_{j=0}^{n} \sum_{k=0}^{n} \sum_{m=0}^{n} \left( \begin{array}{c} r+n \cr r \end{array} \right) \left( \begin{array}{c} s+n \cr s \end{array} \right) \left( \begin{array}{c} t+n \cr t \end{array} \right) \left( \begin{array}{c} u+n \cr u \end{array} \right) \) can be derived, where \( \nu \) is a fraction and
\[
(1) \quad \Gamma(n+1)^{-1} \Gamma(\nu) = \frac{\Gamma(n+1)^{-1} \Gamma(\nu)}{\Gamma(n+1)^{-1} \Gamma(n+1)}.
\]
Thus, from (1), it follows that:
\[
D_{\nu}^{r} \hat{F}^* = \sum_{i=0}^{n} \left[ \begin{array}{c} r+n \cr r \end{array} \right] D_{\nu}^{r} \hat{F}^*[D_{\nu}^{r} \hat{F}^*],
\]
where \( x \) is an independent variable, \( \nu \) is the \( k \)th net input of \( f^* \), and \( a \) is a constant. When \( n \gg r \), \( \nu \) is the domain of definition of \( \nu \) and \( \nu^* \), respectively. \( \hat{F} = \hat{F}^*[\nu, \nu^*] \) is a composite function, \( D_{\nu}^{r} \), \( D_{\nu}^{r} \), and \( D_{\nu}^{r} \) are the integer-order differential operators, and nonnegative integer \( p_i \) satisfies:
\[
\sum_{i=1}^{n} p_i = n \quad \text{and} \quad \sum_{i=1}^{n} p_i = h.
\]
The third summation notations \( \sum_{i=1}^{n} \) in (7) and (8) denote the summation of the corresponding
\[
\left[ \prod_{i=1}^{n} \left[ \frac{1}{p_i} \right] \prod_{i=1}^{n} \left[ \frac{1}{p_i} \right] \right] \left[ \frac{1}{p_i} \right] \prod_{i=1}^{n} \left[ \frac{1}{p_i} \right] \prod_{i=1}^{n} \left[ \frac{1}{p_i} \right]
\]
that satisfy the requirement of (9).

From Fig. 1, we can observe that \( \nu^* = w \nu \nu^* + b^* \). Thus, the net input \( \nu^* \) of \( f^* \) is as follows:
\[

\nu^* = \sum_{i=1}^{n} \nu_{i} \nu_{i}^* + b^*.
\]

From (10), it follows that:
\[
D_{\nu}^{r} \nu_{i} \nu_{i}^* \nu_{i}^* = \beta_{i}^*.
\]

From (11) and (12), it follows that:
\[
D_{\nu}^{r} \nu_{i} \nu_{i}^* \nu_{i}^* = 0.
\]

Using mathematical induction, it can be proven that from (9), we can derive that:
\[
\left\{ \begin{array}{l}
\sum_{i=1}^{n} p_i = 0, \quad \text{when } h = 1, \\
p_i = 1, \quad \text{for } h = 1, \\
p_{i+1} = p_{i+1}, \quad \text{for } i = 1, \ldots, n-1, \text{when } h = 2, \\
p_i = p_i = \cdots = p_0 = 0, \quad \text{when } h = n.
\end{array} \right\
\]

Thus, from (7) and (8), if and only if:
\[
D_{\nu}^{r} \nu_{i} \nu_{i}^* \nu_{i}^* = 0.
\]

From (13) and (14), we can derive that:
\[
\sum_{i=1}^{n} p_i = 0, \quad \text{when } h = 1, \\
p_i = 1, \quad \text{for } h = 1, \\
p_{i+1} = p_{i+1}, \quad \text{for } i = 1, \ldots, n-1, \text{when } h = 2, \\
p_i = p_i = \cdots = p_0 = 0, \quad \text{when } h = n.
\]

Thus, from (11) and (12), it follows that:
\[
D_{\nu}^{r} \nu_{i} \nu_{i}^* \nu_{i}^* = 0.
\]

Using mathematical induction, it can be proven that from (9), we can derive that:
\[
\left\{ \begin{array}{l}
\sum_{i=1}^{n} p_i = 0, \quad \text{when } h = 1, \\
p_i = 1, \quad \text{for } h = 1, \\
p_{i+1} = p_{i+1}, \quad \text{for } i = 1, \ldots, n-1, \text{when } h = 2, \\
p_i = p_i = \cdots = p_0 = 0, \quad \text{when } h = n.
\end{array} \right\}
\]

Thus, from (11) and (12), it follows that:
\[
D_{\nu}^{r} \nu_{i} \nu_{i}^* \nu_{i}^* = 0.
\]

Using mathematical induction, it can be proven that from (9), we can derive that:
\[
\left\{ \begin{array}{l}
\sum_{i=1}^{n} p_i = 0, \quad \text{when } h = 1, \\
p_i = 1, \quad \text{for } h = 1, \\
p_{i+1} = p_{i+1}, \quad \text{for } i = 1, \ldots, n-1, \text{when } h = 2, \\
p_i = p_i = \cdots = p_0 = 0, \quad \text{when } h = n.
\end{array} \right\}
\]

Thus, from (11) and (12), it follows that:
\[
D_{\nu}^{r} \nu_{i} \nu_{i}^* \nu_{i}^* = 0.
\]

Using mathematical induction, it can be proven that from (9), we can derive that:
\[
\left\{ \begin{array}{l}
\sum_{i=1}^{n} p_i = 0, \quad \text{when } h = 1, \\
p_i = 1, \quad \text{for } h = 1, \\
p_{i+1} = p_{i+1}, \quad \text{for } i = 1, \ldots, n-1, \text{when } h = 2, \\
p_i = p_i = \cdots = p_0 = 0, \quad \text{when } h = n.
\end{array} \right\}
\]
Thus, from (22), (23), (25), and (26), it follows that:

\[
\begin{align*}
\hat{w}_m^\mu(k+1) &= w_m^\mu(k) - \mu \left[ D_{\hat{v}_m} \hat{F}(k) \right], \\
\hat{b}_m^\mu(k+1) &= b_m^\mu(k) - \mu \left[ D_{\hat{v}_m} \hat{F}(k) \right].
\end{align*}
\]  
(29)

Equations (20) and (21) indicate that although the first four terms of \( D_{\hat{v}_m} \hat{F}[g^\mu(w_m^\mu, b_m^\mu)] \) and \( D_{\hat{v}_m} \hat{F}[g^\mu(w_m^\mu, b_m^\mu)] \) are only taken into account, \( D_{\hat{v}_m} \hat{F}[g^\mu(w_0^\mu, b_0^\mu)] \) and \( D_{\hat{v}_m} \hat{F}[g^\mu(w_0^\mu, b_0^\mu)] \) are essentially the approximate fractional differentials of the square error \( \hat{F} \) of a BPNN, which still preserve the strengths of fractional calculus such as long-term memory, nonlocality, and weak singularity. Therefore, in a similar way to the FSDM, from (2), (20), and (21), the modified FSDM for the family of the BPNNs can be implemented as follows:

\[
\begin{align*}
\hat{w}_m^\mu(k+1) &= w_m^\mu(k) - \mu \left[ D_{\hat{v}_m} \hat{F}(k) \right], \\
\hat{b}_m^\mu(k+1) &= b_m^\mu(k) - \mu \left[ D_{\hat{v}_m} \hat{F}(k) \right].
\end{align*}
\]  
(30)

where \( \hat{w}_m^\mu, \hat{b}_m^\mu \) is the weight between the \( m^{th} \) output of the \( (m-1)^{th} \) layer and the \( i^{th} \) input of the \( m^{th} \) layer, and \( \mu \) is the learning rate of a modified FSDM (a positive small number). Note that in (22) and (23), even if \( \hat{D}_{\hat{v}_m} \hat{F}(k)=0 \) and \( \hat{D}_{\hat{v}_m} \hat{F}(k)=0 \), but \( \hat{F}(k)=0 \), the iterative search processes should be constrained to keep going on. When \( \hat{D}_{\hat{v}_m} \hat{F}(k)=0 \) and \( \hat{D}_{\hat{v}_m} \hat{F}(k)=0 \), but \( \hat{F}(k)=0 \), the corresponding point \( (w_0^\mu, b_0^\mu) \) is merely a saddle point rather than a real fractional-order optimal minimum point of a modified FSDM. Equations (22) and (23) can be expressed in vector form.

\[
\begin{align*}
\hat{w}_m^\mu(k+1) &= w_m^\mu(k) - \mu \left[ D_{\hat{v}_m} \hat{F}(k) \right], \\
\hat{b}_m^\mu(k+1) &= b_m^\mu(k) - \mu \left[ D_{\hat{v}_m} \hat{F}(k) \right].
\end{align*}
\]  
(31)

In this subsection, the mathematical proof of the fractional-order global optimal convergence of a modified FSDM based FBPN is presented.

Assume that the square error \( \hat{F}(k) \) of a modified FSDM based FBPN is a smooth convex function with at least one of the equivalent fractional-order optimal extreme points \( (w_0^\mu, b_0^\mu) \). Each iterative step of the iterative search processes of a modified FSDM based FBPN is formulated as (22), (23), (27) and (28). The fingerprint of a fractional-order optimal minimum point is its corresponding point \( \hat{F}(k) \) where \( \hat{D}_{\hat{v}_m} \hat{F}(k) \) and \( \hat{D}_{\hat{v}_m} \hat{F}(k) \) are equal to zero. Therefore, we can obtain the following lemma.

**Lemma 1:** If the number of neurons and number of hidden layers of a modified FSDM based FBPN are such that at least one fractional-order minimum point of the square error \( \hat{F}(k) \) exists, even if \( \hat{D}_{\hat{v}_m} \hat{F}(k)=0 \) and \( \hat{D}_{\hat{v}_m} \hat{F}(k)=0 \), but \( \hat{F}(k)=0 \), the iterative search processes are constrained to keep going on, then the modified FSDM described in (22), (23), (27) and (28) can guarantee this FBPN converge to a fractional-order optimal minimum point \( (w_0^\mu, b_0^\mu) \), a global minimum point, of the square error \( \hat{F}(k) \).

**Proof:** Equations (22), (23), and (25)–(28) indicate that on a fractional-order optimal minimum point \( (w_0^\mu, b_0^\mu) \) of the square error \( \hat{F}(k) \), one can obtain \( \hat{D}_{\hat{v}_m} \hat{F}(k)=0 \) and \( \hat{D}_{\hat{v}_m} \hat{F}(k)=0 \). Thus, the criteria for the convergence of (27) and (28) are:

\[
\begin{align*}
\hat{w}_m^\mu(k+1) &= w_m^\mu(k) - \mu \left[ D_{\hat{v}_m} \hat{F}(k) \right], \\
\hat{b}_m^\mu(k+1) &= b_m^\mu(k) - \mu \left[ D_{\hat{v}_m} \hat{F}(k) \right].
\end{align*}
\]  
(32)

Note that (27) and (28), and Fig. 1, indicate that by employing the modified FSDM to achieve the training process for the family of the BPNNs, a BPNN is indeed a modified FSDM based FBPN. The architecture of a modified FSDM based FBPN is identical to that of a traditional first-order BPNN; however, the reverse incremental searches of a modified FSDM based FBPN are in the negative direction of the approximate \( n \)th fractional derivatives of quadratic energy norm \( \hat{F}(k) \).

As we know, the classic first-order BA for the approximate mean squared error of a first-order BPNN is as follows [1]–[6]:

\[
\hat{w}_m^\mu(k+1) = w_m^\mu(k) - \mu \left[ D_{\hat{v}_m} \hat{F}(k) \right], \\
\hat{b}_m^\mu(k+1) = b_m^\mu(k) - \mu \left[ D_{\hat{v}_m} \hat{F}(k) \right].
\]  
(33)

The reverse incremental search of a classic first-order BPNN is in the negative direction of the first-order derivative of the square error \( \hat{f}(k) \). Comparing (27) and (28) with (29) and (30), we can observe that when \( v=1 \), a modified FSDM based FBPN converts to be a classic first-order BPNN. The first-order FBNN is a special case of a modified FSDM based FBPN. The classic first-order optimal minimum point is a particular case of the fractional-order minimum one [40].

**C. Fractional-order global optimal convergence of modified FSDM based FBPN**

In this subsection, the mathematical proof of the fractional-order global optimal convergence of a modified FSDM based FBPN is presented.
The layer, the last layer, from (3) and (10), we can derive that:

\[ D_{ik}^{m} \beta_{m} = w_{m}^{n} \left[ D_{ik}^{m} f^{m}(g_{k}^{m}) \right] = w_{m}^{n} \left[ D_{ik}^{m} f^{m} \left( \sum_{j} w_{j} \beta_{j} + b_{j} \right) \right] \]  

(32)

Thus, from (32), it follows that:

\[ \rho^{n}(k) = D_{ik}^{m} \beta_{m}^{n} \left[ D_{ik}^{m} f^{m}(g_{k}^{m}) \right] = \left( D_{ik}^{m} f^{m} \left( \sum_{j} w_{j} \beta_{j} + b_{j} \right) \right) \beta_{m}^{n} \]  

(33)

which is closely related to the activation function of the last layer of a modified FSDM based FBPN. As a similar to (33), (24) and Fig. 1, we can derive that:

\[ \rho^{n}(k) = \left( D_{ik}^{m} f^{m}(g_{k}^{m}) \right) \beta_{m}^{n} \]  

(34)

Equation (34) indicates that the sensibility back propagates from the last layer to the first layer of the modified FSDM based FBPN. In particular, with regard to the modified layer of the last layer of a FBPNN, from (3), (5), (24), and (24), we can derive that:

\[ \beta_{m}^{n} = D_{ik}^{m} \beta_{m}^{n} = \left( D_{ik}^{m} f^{m} \left( \sum_{j} w_{j} \beta_{j} + b_{j} \right) \right) \beta_{m}^{n} \]  

(35)

Substitute (36) into (35), we obtain:

\[ \rho^{n}(k) = -2[g_{j} - \beta] \beta_{m}^{n} = -2q_{j} \beta \beta_{m}^{n} \]  

(37)

From (24), (35), and (37), we have:

\[ \rho^{n}(k) = D_{ik}^{m} \beta_{m}^{n} \left( h_{k}^{m} \right) = -2q_{j} \beta \left[ D_{ik}^{m} f^{m} \left( \sum_{j} w_{j} \beta_{j} + b_{j} \right) \right] \]  

(38)

\[ \rho^{n}(k) = D_{ik}^{m} \beta_{m}^{n} \left( h_{k}^{m} \right) = -2q_{j} \beta \left[ D_{ik}^{m} f^{m} \left( \sum_{j} w_{j} \beta_{j} + b_{j} \right) \right] \]  

(39)

Equations (37)–(39) are the initial point values of the backpropagation recurrence relations, which can be expressed by (34).

In addition, if \( D_{ik}^{m} f(k) = 0 \) and \( D_{ik}^{m} f(k) = 0 \), but \( f(k) \neq 0 \), as long as the iterative search processes are constrained to keep going on, then, when the iterative search processes converge to a fractional-order optimal minimum point \( \left( \beta_{u}^{m}, h_{u}^{m} \right) \) of the square error \( F(k) \), with respect to the modified layer of a modified FSDM based FBPN, to enable (31) to be set up, from (37), (38), and (39), a sufficient and necessary condition can be given as:

\[ F(k) + \sum_{j=1}^{n} \left( q_{j} - \beta \right) = 0 \]  

(40)

where \( \beta = \beta_{u}^{m} \) and \( i = 1, 2, \ldots, m \). To enable (40) to be set up, a sufficient and necessary condition can be given as:

\[ \begin{align*}
  q_{j} - \beta & = 0 \\
  D_{ik}^{m} f(k) & = 0
\end{align*} \]  

(41)

Further, when (41) is set up, we have:

\[ \begin{align*}
  \sum_{j=1}^{n} \left( q_{j} - \beta \right) &= 0 \\
  -2q_{j} \beta & = 0
\end{align*} \]  

(42)

Therefore, when the iterative search processes converge to a fractional-order optimal minimum point \( \left( \beta_{u}^{m}, h_{u}^{m} \right) \) of the square error \( F(k) \) of a modified FSDM based FBPN, from (5), (35), and (42), we obtain:

\[ \begin{align*}
  F(k) + \sum_{j=1}^{n} \left( q_{j} - \beta \right) &= 0 \\
  D_{ik}^{m} f(k) & = -2q_{j} \beta
\end{align*} \]  

(43)

Thus, from (34), when (43) is set up, we have:

\[ \begin{align*}
  F(k) & = 0 \\
  D_{ik}^{m} f(k) & = 0
\end{align*} \]  

(44)

where \( m = 0, 1, \ldots, M - 1 \), \( D_{ik}^{m} f(k) = \beta_{u}^{m} h_{u}^{m} = \left[ D_{ik}^{m} f(k) \right] \beta_{u}^{m} \), and \( D_{ik}^{m} f(k) = \beta_{u}^{m} h_{u}^{m} \left[ D_{ik}^{m} f(k) \right] \beta_{u}^{m} \). Note that in (41), (43), and (44), for \( f(k) \geq 0 \), thus when \( f(k) = 0 \), \( \beta \) is constant. Thus, if \( f(k) = 0 \), we obtain \( D_{ik}^{m} f(k) = 0 \) and \( \beta_{u}^{m} h_{u}^{m} = 0 \). Thus, in fact, in (41), (43) and (44), the second formula can be derived from the first one, but not vice versa. As we know, when and only when (43) and (44) are set up, the point of convergence of the iterative search processes (the final convergence result) is the same as the global minimum point of the square error \( F(k) \) of a FBNN, namely, the iterative search algorithms in (27) and (28) can be guaranteed to converge to a fractional-order optimal minimum point \( \left( \beta_{u}^{m}, h_{u}^{m} \right) \), a global minimum point, of the square error \( F(k) \) of a modified FSDM based FBPN. This completes the proof.

Lemma 1 indicates that at first, in (43) and (44), \( D_{ik}^{m} f(k) = \rho_{u}^{m} h_{u}^{m} \) and \( D_{ik}^{m} f(k) = \rho_{u}^{m} h_{u}^{m} \) merely consider the local characteristics of the square error \( F(k) \) of a classic first-order BPN. Thus, a classic first-order BPN is likely to converge to a local extreme point of its square error \( F(k) \). Assume that the iterative search process of a classic first-order BPN converges to a local extreme point \( \left( \beta_{u}^{m}, h_{u}^{m} \right) \). Thus, on this local extreme point \( \left( \beta_{u}^{m}, h_{u}^{m} \right) \), one can obtain:
Formulated in (3) is a modified FSDM based FBPNN, which is also highly dependent on the number of neurons and that of hidden layers. The aforementioned properties highlighted for a modified FSDM based FBPNN regarding the fractional-order multi-scale global optimization search assume the sizing of a modified FSDM based FBPNN is sufficient such that at least a fractional-order minimum point (a global minimum point) exists. With regards to an undersized modified FSDM based FBPNN, it does not exit a zero square error \( \hat{f}(k) \). Thus, it is possible that the domain of attraction of the fractional-order optimal minimum point of an undersized modified FSDM based FBPNN does not contain the attractor \( \{\omega_i^*, \kappa_i^*\} \), or this domain of attraction is not included in the state space of an undersized modified FSDM based FBPNN at all. Assume that there is a local extreme point \( \{\omega_i^*, \kappa_i^*\} \) different from the fractional-order optimal minimum point \( \{\omega_i^*, \kappa_i^*\} \) of the square error \( \hat{f}(k) \). Thus, on this local extreme point \( \{\omega_i^*, \kappa_i^*\} \), substitution of (45) into (27) and (28) results in:

\[
\begin{align*}
\|\omega_i^*(k+1) - \omega_i^*(k)\| &= \|\omega_i^*(k) - \omega_i^*\| \neq 0 \\
\|\kappa_i^*(k+1) - \kappa_i^*(k)\| &= \|\kappa_i^*(k) - \kappa_i^*\| \neq 0
\end{align*}
\]

Equation (46) indicates that the iterative search algorithms in (27) and (28) on a local extreme point \( \{\omega_i^*, \kappa_i^*\} \) cannot be ultimately terminated by themselves. On the one hand, if the domain of attraction of point \( \{\omega_i^*, \kappa_i^*\} \) is not entire and point \( \{\omega_i^*, \kappa_i^*\} \) is in only this domain of attraction, the iterative search process of an undersized modified FSDM based FBPNN should eventually somewhat oscillate around point \( \{\omega_i^*, \kappa_i^*\} \) and attempt to identify point \( \{\omega_i^*, \kappa_i^*\} \). Otherwise, it should deviate from point \( \{\omega_i^*, \kappa_i^*\} \) to further determine point \( \{\omega_i^*, \kappa_i^*\} \). Conversely, if the domain of attraction of point \( \{\omega_i^*, \kappa_i^*\} \) is not included in the state space at all, the iterative search process of an undersized modified FSDM based FBPNN should deviate from point \( \{\omega_i^*, \kappa_i^*\} \) to make an continual attempt to search for point \( \{\omega_i^*, \kappa_i^*\} \).

Secondly, regarding a multilayer perceptron, Cybenko first mathematically proved that arbitrary decision regions can be arbitrarily well approximated by continuous feedforward neural networks with only a single hidden layer and any continuous sigmoidal nonlinearity [46]. Sontag derived a general result demonstrating that nonlinear control systems can be stabilized using two hidden layers; however, not in general using only a single hidden layer in a classic first-order BPN [47]. Barron examined how the approximation error is related to the number of neurons in a classic first-order BPN [48]. In a similar manner to a first-order BPN, the approximation properties related to the number of neurons of the hidden layer in a modified FSDM based FBPNN must be established. For the convenience of discussion, without loss of generality, assume that the actual output function \( f_p'(\theta_n) \) on \( \mathbb{R}^m \) formulated in (3) is approximated by a modified FSDM based FBPNN with a single layer of sigmoidal units, given as:

\[
f_p'(\theta_n) = W_{p_m} \varphi'(\psi_i^{\alpha})
\]

where \( \psi_i^{\alpha} \geq 1 \) and

\[
\begin{align*}
\hat{f}(k) &= 0 \\
\omega_i^*(k) &= D_{\omega_i}(k) = \hat{f}(k) = 0
\end{align*}
\]

However, in (25) and (26),

\[
\begin{align*}
D_{\omega_i}(k) &= \frac{\sum_{c=1}^{v} |\sum_{n=1}^{c} \sum_{l=1}^{m} \omega_{ji} \cdot w_{in} |}{(1-v)} \\
D_{\omega_i}(k) &= \frac{\sum_{c=1}^{v} |\sum_{n=1}^{c} \sum_{l=1}^{m} \omega_{ji} \cdot w_{in} |}{(1-v)} \end{align*}
\]

consider the nonlocal characteristics and the weak singularity of the square error \( \hat{f}(k) \) of a modified FSDM based FBPNN. Thus, a modified FSDM based FBPNN can be guaranteed to converge to a global minimum point, of its square error \( \hat{f}(k) \). The more efficient optimal searching capability of the fractional-order multi-scale global optimization of a modified FSDM based FBPNN to determine the global optimal solution is the major advantage being superior to a classic first-order BPN. Secondly, if the square error \( \hat{f}(k) \) is mixed with white noise, it should increase the convergence time of the reverse incremental search of a modified FSDM based FBPNN, which could cause (41) to be not set up. In this case, (41) could be invalid and hence the FBPNN may not converge to the fractional-order optimal point. Thirdly, the computational complexity of an algorithm can be typically measured by the number of its multiplications and additions, and related memory space, thus the computational complexity of a BPN is linear with \( \psi = [\omega_i]_{i=1}^{m} \) and \( \psi = [\kappa_i]_{i=1}^{m} \), which is in direct proportion to \( \Omega_{\psi} = [\omega_{ij}]_{i=1}^{m} \) (\( \Omega_{\psi} \) is the number of free parameters of a matrix). Therefore, for \( n = 1, 2, 3 \) in (27) and (28), the computational complexity of a modified FSDM based FBPNN is in direct proportion to \( \Omega_{\psi}^{n} \). Therefore, with the same number of neurons, the computational complexity of a modified FSDM based FBPNN is 4 times greater than that of a classic first-order BPN. In particular, Equations (27), (28), and (41) show that for the final convergence result of a modified FSDM based FBPNN with \( n = 1 \) is the same as that with \( n = 2, 3 \). Thus, to further simplify the calculation of (27) and (28), without loss of generality, in the actual computations of a modified FSDM based FBPNN, we can only set \( n = 1 \). Then, in this case, with the same number of neurons, the computational complexity of a modified FSDM based FBPNN is 2 times greater than that of a classic first-order BPN. Fourthly, in general, the fractional-order external extreme value of a normalized quadratic energy norm determined by the fractional-order partial derivatives is not equal to its integer-order one [40]. However, (3), (5), (41), and (44) indicate that if the global minimum value of the square error \( \hat{f}(k) \) is equal to zero, the fractional-order optimal minimum point \( \{\omega_i^*, \kappa_i^*\} \) determined by the approximate fractional-order partial derivatives expressed by (25) and (26) is identical to the global minimum value of the square error \( \hat{f}(k) \).

D. Assumption of structure of modified FSDM based FBPNN

In this subsection, an assumption of the structure of a modified FSDM based FBPNN is made.

At first, in a similar manner to a classic first-order BPN, a modified FSDM based FBPNN behaviour is also highly
denotes the number of neurons of the hidden layer, and \( f \) is an arbitrary fixed sigmoid function. Assume that \( \mathbf{p}_{0,\gamma}(\mathbf{p}_{\alpha,\gamma}) \) is a target output function of a modified FSMD based FBPNN. If \( \mathbf{p}_{0,\gamma}(\mathbf{p}_{\alpha,\gamma}) \) is a causal signal with fractional primitives zero, from (1), we can derive the Fourier transform of the \( \nu \)-order derivative of \( \mathbf{p}_{0,\gamma}(\mathbf{p}_{\alpha,\gamma}) \): given as:

\[
\mathbf{F}^{(\nu)}(\mathbf{p}_{0,\gamma}(\mathbf{p}_{\alpha,\gamma})) = (c_{\alpha,\gamma}, \mathbf{p}_{0,\gamma}(\mathbf{p}_{\alpha,\gamma})), \tag{48}
\]

where \( \mathbf{F}^{(\nu)} \) denotes the Fourier transform, \( \zeta \) denotes an imaginary unit, \( \mathbf{p}_{0,\gamma}(\mathbf{p}_{\alpha,\gamma}) \) denotes angular frequency, and \( \mathbf{p}_{0,\gamma}(\mathbf{p}_{\alpha,\gamma}) = \sum_{\mathbf{p}_{\alpha,\gamma}} \mathbf{p}_{0,\gamma}(\mathbf{p}_{\alpha,\gamma}) \) is the Fourier transform of \( \mathbf{p}_{0,\gamma}(\mathbf{p}_{\alpha,\gamma}) \). Thus, from (48), the smoothness property of \( \mathbf{p}_{0,\gamma}(\mathbf{p}_{\alpha,\gamma}) \) can be measured by the integrability of \( \mathbf{F}^{(\nu)}(\mathbf{p}_{0,\gamma}(\mathbf{p}_{\alpha,\gamma})) \) on \( \mathfrak{m}^{\alpha,\gamma} \), given as:

\[
\mathcal{C}_{\gamma} = \mathcal{F}^{(\nu)}(\mathbf{p}_{0,\gamma}(\mathbf{p}_{\alpha,\gamma})) = \int \mathfrak{m}^{\alpha,\gamma} \mathbf{p}_{0,\gamma}(\mathbf{p}_{\alpha,\gamma}) \mathfrak{m}^{\alpha,\gamma}, \tag{49}
\]

where \( \mathfrak{m}^{\alpha,\gamma} \) is the integral (49) and \( \cdot \) denotes the inner product. Assume that the approximation error of a modified FSMD based FBPNN can be measured by the integrated squared error regarding a probability \( \rho \) on the hyper-ball \( \mathcal{B} = \{ r \mathfrak{m}^{\alpha,\gamma} | r \leq r \} \) of radius \( r = \sup_{\rho} \mathfrak{m}^{\alpha,\gamma} > 0 \). Therefore, we have the following lemma.

**Lemma 2:** If a target output function \( \mathbf{p}_{0,\gamma}(\mathbf{p}_{\alpha,\gamma}) \) with \( \mathcal{C}_{\gamma} \geq 0 \) is in the closure of the convex hull of a set \( \mathfrak{g} \) in a Hilbert space, with \( \| \mathfrak{g} \| < b \) for each \( \mathfrak{g} \in \mathfrak{g} \) and \( \mathcal{C}_{\gamma} = \| \mathfrak{g} \| \geq 0 \), there should be an actual output function \( \mathbf{p}_{\gamma}(\mathbf{p}_{\alpha,\gamma}) \) of a modified FSMD based FBPNN formulated in (47) in the convex hull of \( \gamma \geq 1 \) points in \( \mathfrak{g} \), such that:

\[
\| \mathbf{p}_{\gamma}(\mathbf{p}_{\alpha,\gamma}) - \mathbf{p}_{0,\gamma}(\mathbf{p}_{\alpha,\gamma}) \| _{\rho} \leq \mathcal{C}_{\gamma}, \tag{50}
\]

where \( \| \| \) denotes the norm of the Hilbert space. If \( \mathbf{p}_{\gamma}(\mathbf{p}_{\alpha,\gamma}) \) is observed at sites \( \rho = \rho_{\gamma} \) restricted to \( \mathcal{B} \) that obey the uniform distribution, (50) provides a restricted condition on the approximation error, given as:

\[
\frac{1}{S} \sum_{\gamma} \| \mathbf{p}_{\gamma}(\mathbf{p}_{\alpha,\gamma}) - \mathbf{p}_{0,\gamma}(\mathbf{p}_{\alpha,\gamma}) \| _{\rho} \leq \mathcal{C}_{\gamma}, \tag{51}
\]

where \( S \) is the sample size.

**Proof:** We prove Lemma 2 based on the law of large numbers. Assume that \( \mathbf{p}_{\gamma}(\mathbf{p}_{\alpha,\gamma}) \) is a point in the convex hull of \( \mathfrak{g} \) being extremely close to \( \mathbf{p}_{\gamma}(\mathbf{p}_{\alpha,\gamma}) \), with \( \| \mathbf{p}_{\gamma}(\mathbf{p}_{\alpha,\gamma}) - \mathbf{p}_{\gamma}(\mathbf{p}_{\alpha,\gamma}) \| _{\rho} \leq \delta \) and \( \delta \geq 0 \). \( \mathfrak{g} \) is stochastically drawn from the set \( \mathfrak{g} = \sum_{\gamma} \mathfrak{g} \) with probability \( \rho_{\gamma} \leq 0 \). Then, if \( m \) is sufficiently large, \( \mathbf{p}_{\gamma}(\mathbf{p}_{\alpha,\gamma}) = \frac{1}{m} \sum_{\gamma} \mathfrak{g} \) is independently drawn from the same distribution as \( \mathfrak{g} \). For the uniform distribution, \( \rho = \rho_{\gamma} = 1/\rho \). If \( \gamma \) is sufficiently large, \( \mathbf{p}_{\gamma}(\mathbf{p}_{\alpha,\gamma}) = 1/\rho \sum_{\gamma} \mathfrak{g} \) is a sample average. Thus,

\[
E[\mathbf{p}_{\gamma}(\mathbf{p}_{\alpha,\gamma})] = \mathbf{p}_{0,\gamma}(\mathbf{p}_{\alpha,\gamma}), \tag{52}
\]

where \( E[.] \) denotes the mathematical expectation. Therefore,

\[
E[\mathbf{p}_{\gamma}(\mathbf{p}_{\alpha,\gamma}) - \mathbf{p}_{\gamma}(\mathbf{p}_{\alpha,\gamma})] = \frac{1}{S} \sum_{\gamma} E[\mathbf{p}_{\gamma}(\mathbf{p}_{\alpha,\gamma}) - \mathbf{p}_{\gamma}(\mathbf{p}_{\alpha,\gamma})], \tag{53}
\]

which equals \( \mathcal{C}_{\gamma} \). For the mathematical expectation to be bounded in this manner, there must be \( \mathcal{C}_{\gamma} \leq 0 \), for which \( \mathcal{C}_{\gamma} \leq 0 \). Because

\[
\mathcal{C}_{\gamma} = \frac{1}{S} \sum_{\gamma} \| \mathbf{p}_{\gamma}(\mathbf{p}_{\alpha,\gamma}) - \mathbf{p}_{\gamma}(\mathbf{p}_{\alpha,\gamma}) \| _{\rho} \leq \mathcal{C}_{\gamma}, \tag{54}
\]

the proof of Lemma 2 is completed by the choice of a sufficiently small \( \delta \).

Furthermore, let \( \mathcal{C}_{\gamma} = (C_{\gamma})_{\gamma} \) and add a statistically estimated regularization item \( -\frac{R_{\gamma}}{S} \log S \) to the restricted condition on the approximation error of a modified FSMD based FBPNN formulated in (51), given as:

\[
\frac{1}{S} \sum_{\gamma} \| \mathbf{p}_{\gamma}(\mathbf{p}_{\alpha,\gamma}) - \mathbf{p}_{\gamma}(\mathbf{p}_{\alpha,\gamma}) \| _{\rho} \leq \frac{C_{\gamma}}{\rho} - \frac{R_{\gamma}}{S} \log S, \tag{55}
\]

where \( R_{\gamma} \) is the number of input nodes. Equation (52) indicates that if the approximation error \( \frac{1}{S} \sum_{\gamma} \| \mathbf{p}_{\gamma}(\mathbf{p}_{\alpha,\gamma}) - \mathbf{p}_{\gamma}(\mathbf{p}_{\alpha,\gamma}) \| _{\rho} = 0 \), the number of neurons of the hidden layer in a modified FSMD based FBPNN can be derived as:

\[
\gamma = C_{\gamma} \frac{R_{\gamma}}{S \log S}, \tag{56}
\]

Equation (53) indicates that when the number of neurons of the hidden layer \( \gamma = C_{\gamma} \frac{R_{\gamma}}{S \log S} \), the sizing of a modified FSMD based FBPNN is sufficient such that at least one fractional-order minimum exists.

E. Fractional-order multi-scale global optimization of modified FSMD based FBPNN

In this subsection, the fractional-order multi-scale global optimization of a modified FSMD based FBPNN is analysed. Equations (25)–(28) indicate that the variation of the fractional-order \( \nu \) of the \( \bar{\mathbf{F}} \) and \( \mathbf{F} \) can actually non-linearly influence the learning process of a modified FSMD based FBPNN to a certain degree. We restrict the square error \( \| \hat{e}(k) \| \) to be a non-linear increasing function of the fractional-order \( \nu(k) \). Further, to simplify the calculation, we restrict \( 0 < \nu < 2 \).

At first, (5), (25), and (26) indicate that on a given local extreme point and in its neighbourhood, we have \( 0 < \nu < 1 \) and \( \bar{\mathbf{F}} \) is approximately a sufficiently small positive scalar and a sufficiently large positive scalar, respectively. To escape from a local extreme point and its neighbourhood, the iterative search process of a modified FSMD based FBPNN should have climbing capacity on \( \nu(k) \) by itself, i.e., \( \bar{\mathbf{F}} < \mathbf{F} \) should be in nearly opposite directions of \( \bar{\mathbf{F}} \) and \( \mathbf{F} \), respectively. On the one hand, if \( \bar{\mathbf{F}} > 0 \) or \( \mathbf{F} > 0 \), we should restrict \( \bar{\mathbf{F}} < \mathbf{F} \) or \( \bar{\mathbf{F}} > \mathbf{F} \) correspondingly. Thus, from (25)–(28), we can derive that:
Further, if $0<\sigma_2<\sigma_1$, to enable (56) and (57) to be set up, a necessary condition is given as:

$$
\tilde{F}(k) > -\left(1 - \frac{1}{\sigma_1} \right) \sum_{n=0}^{\infty} \left( \frac{1}{\sigma_1} \right)^n \sum_{n=0}^{\infty} \left( \frac{1}{\sigma_1} \right)^n \rho_0(k, \beta_0^n) > 0.
$$

In (68) and (70), $\sigma_2 = \min\{\sigma(\nu, \sigma(\nu_1))\}$ and $\sigma_1 = \max\{\sigma(\nu, \sigma(\nu_1))\}$. In (69) and (71), $\sigma_2 = \max\{\sigma(\nu_1, \sigma(\nu_1))\}$ and $\sigma_1 = \min\{\sigma(\nu_1, \sigma(\nu_1))\}$. When $\tilde{F} > 0$, we should restrict $\tilde{F} > 0$ corresponding to (56) and (58) in this case, the following is true:

$$
\tilde{F}(k) > -\left(1 - \frac{1}{\sigma_1} \right) \sum_{n=0}^{\infty} \left( \frac{1}{\sigma_1} \right)^n \sum_{n=0}^{\infty} \left( \frac{1}{\sigma_1} \right)^n \rho_0(k, \beta_0^n) > 0.
$$

Further, if $0<\sigma_2<\sigma_1$, to enable (56) and (57) to be set up, a necessary condition is given as:

$$
\tilde{F}(k) > -\left(1 - \frac{1}{\sigma_1} \right) \sum_{n=0}^{\infty} \left( \frac{1}{\sigma_1} \right)^n \sum_{n=0}^{\infty} \left( \frac{1}{\sigma_1} \right)^n \rho_0(k, \beta_0^n) > 0.
$$

Thus, if $0<\nu_1<1$, from (58) and (59), according to the properties of inequality, (56) and (57) can be further simplified as:

$$
\tilde{F}(k) > \left(1 - \frac{1}{\sigma_1} \right) \sum_{n=0}^{\infty} \left( \frac{1}{\sigma_1} \right)^n \sum_{n=0}^{\infty} \left( \frac{1}{\sigma_1} \right)^n \rho_0(k, \beta_0^n) > 0.
$$

In (60) and (61), one can further derive that:

$$
0 < \nu_1 = \left[1 - \left(1 - \frac{1}{\sigma_1} \right) \sum_{n=0}^{\infty} \left( \frac{1}{\sigma_1} \right)^n \sum_{n=0}^{\infty} \left( \frac{1}{\sigma_1} \right)^n \rho_0(k, \beta_0^n) \right] > 1 < v_1 < 1.
$$

In (62), $\sigma_1 = \min\{\sigma(\nu), \sigma(\nu_1)\}$ and $\sigma_2 = \max\{\sigma(\nu), \sigma(\nu_1)\}$. In (63), $\sigma_1 = \min\{\sigma(\nu), \sigma(\nu_1)\}$ and $\sigma_2 = \max\{\sigma(\nu), \sigma(\nu_1)\}$. In addition, if we restrict $1<\nu_1<2$ in this case, we have $-0.5 < \nu_1 < 1$. Thus, if $1<\nu_1<2$, (54), (55) (56) can be further simplified as:

$$
\tilde{F}(k) = \tilde{F}(k) \tilde{F}(k) > -\left(1 - \frac{1}{\sigma_1} \right) \sum_{n=0}^{\infty} \left( \frac{1}{\sigma_1} \right)^n \sum_{n=0}^{\infty} \left( \frac{1}{\sigma_1} \right)^n \rho_0(k, \beta_0^n) > 0.
$$

Further, if $1<\nu_1<2$, (54), (55), and (56) also obtain its maximum value when $n=1$. Thus, if $1<\nu_1<2$, from (66) and (67), (64) and (66) can be further simplified as:

$$
\tilde{F}(k) = \tilde{F}(k) \tilde{F}(k) > -\left(1 - \frac{1}{\sigma_1} \right) \sum_{n=0}^{\infty} \left( \frac{1}{\sigma_1} \right)^n \sum_{n=0}^{\infty} \left( \frac{1}{\sigma_1} \right)^n \rho_0(k, \beta_0^n) > 0.
$$

From (68) and (69), one can further derive that:

$$
\tilde{F}(k) > -\left(1 - \frac{1}{\sigma_1} \right) \sum_{n=0}^{\infty} \left( \frac{1}{\sigma_1} \right)^n \sum_{n=0}^{\infty} \left( \frac{1}{\sigma_1} \right)^n \rho_0(k, \beta_0^n) > 0.
$$

Note that at first, when $0<\sigma_1<\sigma_2<\sigma_3$, $\tilde{F}$ and $\tilde{F}$ are restricted in the near opposite directions of $\tilde{F}$ and $\tilde{F}$, respectively. As we know, the first-order gradient of a point is in the direction of the fastest growth of the scalar field and its module value is the maximum rate of change. In the numerical implementation, the directions of $\tilde{F}$ and $\tilde{F}$ are those of their maximum module values, respectively. For example, on the two-dimensional discrete plane, there are eight directional derivatives (their interval is 45 degrees) of a point $(w_0, b_0)$ in its
neighbourhood. We select the maximum directional derivative of a point \((x_i^*, y_i^*)\) to be \(\Delta x, \Delta y\) and \(\Delta x, \Delta y\). Secondly, to enhance the multi-scale searching capability, we should also set the learning rate, \(\mu(k)\), of a modified FSDM based FBPNN to be an appropriate correlation function with \(\Delta x, \Delta y\) and \(\hat{f}(k)\).

Thirdly, to guarantee to escape the domain of attraction of a given local extreme point, as long as \(\Delta x, \Delta y\) and \(\hat{f}(k)\) begin to climb, we should keep climbing until them get to the top of the hill in the neighbourhood of this local extreme point. On the top of the hill, we should restrict the downward direction to be different from the previous uphill direction of \(\Delta x, \Delta y\) and \(\hat{f}(k)\).

IV. EXPERIMENT AND ANALYSIS

A. Example function approximation of modified FSDM based FBPNN

In this subsection, for the following Examples 1–4, we discuss an example of the function approximation of a modified FSDM based FBPNN.

As we know, multilayer networks can be used to approximate virtually any function if we have a sufficient number of neurons in the hidden layers. We choose a network for a modified FSDM based FBPNN and apply it to a particular problem. Without loss of generality, we illustrate the characteristics with a 1–\(\psi\)–1 modified FSDM based FBPNN, which is displayed in Fig. 2.

In Fig. 2, \(\psi\) denotes the number of neurons in the first layer (hidden layer) of a modified FSDM based FBPNN. The function of the fractional-order \(v\) at the \(k^{th}\) iteration of a modified FSDM based FBPNN as follows:

\[
v(k) = 2 - \frac{\phi}{\rho^v} + \left|\hat{e}(k)\right|
\]

where \(\phi = \frac{1}{|v|} \sum_{i=1}^{m} |e_i(k)|\) is the average error at the \(k^{th}\) iteration, and \(\rho^v = \left(\frac{1}{\rho^v} \sum_{i=1}^{m} \rho_i^v\right)^v\) is the first-order average sensibility in the \(M^v\) layer, the last layer, of a modified FSDM based FBPNN.

From (3), (5), (10), (24), and (78), we can derive the first-order average sensibility in the output layer of the modified FSDM based FBPNN in Fig. 2.

It is observed that the optimal minimum error occurs when \(w_0 = \pm 1\) and \(w_0 = \pm 1\), as indicated by the solid green circle in Fig. 4.

B. Fractional-order multi-scale global optimization of modified FSDM based FBPNN

In this subsection, we analyse the fractional-order multi-scale global optimization of a modified FSDM based FBPNN.

Inspired by the aforementioned mathematical analysis, the fractional-order \(v\) of \(\Delta x, \Delta y\) and \(\hat{f}(k)\) of a modified FSDM based FBPNN can be self-adaptively determined. To simplify the multi-scale searching process in the following experiments, we construct an imperfect adaptive kernel function for the first and second layer are log-sigmoid, which can be expressed as follows:

\[
f(x) = f(x) = \frac{1}{1 + e^{-x}}
\]
The output $\beta^i_n = \beta^i_{n-1}$ of a modified FSDM based FBPNN changes dynamically with the initial condition, weight matrices, bias matrices, and average error $e(k)$. For the convenience of illustration, we analyse the multi-scale adjustment of the adaptive kernel function of $v(k)$ by assuming the output $\beta^i_n = \beta^i_{n-1} = 1 \cdot e(k)$ in (79), which is displayed in Fig. 5.

Fig. 5. Adaptive kernel function of fractional-order $v(k)$.

From Fig. 5 and (79), we can observe that the fractional-order $v(k)$ adaptively varies with the error $e(k)$ during the entire iterative search process of a modified FSDM based FBPNN. As in the previous discussion, if the minimum value of the square error $\hat{f}(k)$ is equal to zero, the fractional order $v(k)$ approaches zero on a fractional-order optimal extreme point (global optimal minimum point) with zero average error $e(k) = 0$.

**Example 1:** In this example, we select an initial condition in the specific zone from a random sample of cases, where the reverse incremental search of a modified FSDM based FBPNN and a classic first-order BPNN can both converge to the global optimal minimum point of the square error $\hat{f}(k)$ at the $k^{th}$ iteration. We set the same parameters, the rate of convergence $\mu = 5.50$ and the number of iterations = 2000, for both a modified FSDM based FBPNN and a first-order BPNN in this simulation experiment; the initial condition is at the point where $w_{1,1} = 4$ and $w_{1,2} = 4$. From (27)–(30), (78), and (79), the iterative search process of a modified FSDM based FBPNN and a first-order BPNN can be represented as displayed in Fig. 6.

In Fig. 6(a), we observe two convergence trajectories of a modified FSDM based FBPNN and a first-order BPNN (batch mode) when only two parameters, $w_{1,1}$ and $w_{1,2}$, are varied. From (27)–(30), the iterative search process of a modified FSDM based FBPNN. As in the previous discussion, if the minimum value of the square error $\hat{f}(k)$ is equal to zero, the fractional order $v(k)$ approaches zero on a fractional-order optimal extreme point (global optimal minimum point) with zero average error $e(k) = 0$.

**Example 2:** In this example, we select an initial condition in the gently sloping zone of the square error $\hat{f}(k)$ from a random sample of cases. We set the same parameters, the rate of convergence $\mu = 5.50$, the number of iterations = 9000, for both a modified FSDM based FBPNN and a first-order BPNN in this simulation experiment; the initial condition is at the point where $w_{1,1} = 5$ and $w_{1,2} = 30$. From (27)–(30), the iterative search process of a modified FSDM based FBPNN and a first-order BPNN can be represented as displayed in Fig. 7.

The square errors $\hat{f}(k)$ of a modified FSDM based FBPNN and a first-order BPNN on an optimal point are equal to zero. The two convergence trajectories of a modified FSDM based FBPNN and a first-order BPNN bypass an initial flat surface and then fall into a gently sloping valley, as observed in Fig. 6(a). From Fig. 6(b), we can observe that a modified FSDM based FBPNN requires a fewer number of iterations than a first-order BPNN to converge to the optimal point. However, as per the aforementioned discussion, with the same number of neurons, the computational complexity of a modified FSDM based FBPNN is 4 times greater than a first-order BPNN in every iterative computation. A modified FSDM based FBPNN requires additional computational time compared to a first-order BPNN. Thus, a first-order BPNN is faster to converge to the optimal point than a modified FSDM based FBPNN. Figure 6(c) indicates that as the value of $\hat{f}(k)$ changes over the convergence trajectory of a modified FSDM based FBPNN, the fractional-order $\beta(k)$ of a modified FSDM based FBPNN varies according to the adaptive kernel function in (79). Thus, the fractional-order $v(k)$ of a modified FSDM based FBPNN in its entire iterative search process is not constant, which demonstrates the fractional-order multi-scale global optimization of a modified FSDM based FBPNN.

In Fig. 7(a), we observe two convergence trajectories of a modified FSDM based FBPNN and a first-order BPNN (batch mode) when only two parameters, $w_{1,1}$ and $w_{1,2}$, are varied. From (27)–(30), the iterative search process of a modified FSDM based FBPNN. As in the previous discussion, if the minimum value of the square error $\hat{f}(k)$ is equal to zero, the fractional order $v(k)$ approaches zero on a fractional-order optimal extreme point (global optimal minimum point) with zero average error $e(k) = 0$.

**Example 2:** In this example, we select an initial condition in the gently sloping zone of the square error $\hat{f}(k)$ from a random sample of cases. We set the same parameters, the rate of convergence $\mu = 5.50$, the number of iterations = 9000, for both a modified FSDM based FBPNN and a first-order BPNN in this simulation experiment; the initial condition is at the point where $w_{1,1} = 5$ and $w_{1,2} = 30$. From (27)–(30), the iterative search process of a modified FSDM based FBPNN and a first-order BPNN can be represented as displayed in Fig. 7.

The square errors $\hat{f}(k)$ of a modified FSDM based FBPNN and a first-order BPNN on an optimal point are equal to zero. The two convergence trajectories of a modified FSDM based FBPNN and a first-order BPNN bypass an initial flat surface and then fall into a gently sloping valley, as observed in Fig. 6(a). From Fig. 6(b), we can observe that a modified FSDM based FBPNN requires a fewer number of iterations than a first-order BPNN to converge to the optimal point. However, as per the aforementioned discussion, with the same number of neurons, the computational complexity of a modified FSDM based FBPNN is 4 times greater than a first-order BPNN in every iterative computation. A modified FSDM based FBPNN requires additional computational time compared to a first-order BPNN. Thus, a first-order BPNN is faster to converge to the optimal point than a modified FSDM based FBPNN. Figure 6(c) indicates that as the value of $\hat{f}(k)$ changes over the convergence trajectory of a modified FSDM based FBPNN, the fractional-order $\beta(k)$ of a modified FSDM based FBPNN varies according to the adaptive kernel function in (79). Thus, the fractional-order $v(k)$ of a modified FSDM based FBPNN in its entire iterative search process is not constant, which demonstrates the fractional-order multi-scale global optimization of a modified FSDM based FBPNN.

**Example 2:** In this example, we select an initial condition in the gently sloping zone of the square error $\hat{f}(k)$ from a random sample of cases. We set the same parameters, the rate of convergence $\mu = 5.50$, the number of iterations = 9000, for both a modified FSDM based FBPNN and a first-order BPNN in this simulation experiment; the initial condition is at the point where $w_{1,1} = 5$ and $w_{1,2} = 30$. From (27)–(30), the iterative search process of a modified FSDM based FBPNN and a first-order BPNN can be represented as displayed in Fig. 7.
In Figs. 7(a) and (b), the convergence trajectory of a first-order BPNN illustrates the manner in which it can converge to a local extreme point \((w_{1}^i = 0.88\) and \(w_{2}^i = 38.66\)). The convergence trajectory of a first-order BPNN is trapped in a valley and diverges from the global optimal solution. However, the convergence trajectory of a modified FSDM based FPBNN illustrates the manner in which it can converge to a global optimal minimum point \((w_{1}^i = 10\) and \(w_{2}^i = 1\)). The convergence trajectory of a modified FSDM based FPBNN passes over a valley and converges to the global optimal solution. Figure 7(c) indicates that when the number of iterations of a first-order BPNN is approximately equal to 6500, the square error \(\hat{e}(k)\) of a first-order BPNN reduces to a nonzero minimum, which is an optimal minimum point (88.01). Illustrates the manner in which it can converge to a global extreme point. Figure 7(c) also indicates that when the number of iterations of a modified FSDM based FPBNN increases from one to ten, its square error \(\hat{e}(k)\) decreases sharply from \(3.4 \times 10^{-3}\) to \(0.35 \times 10^{-3}\). When the number of iterations of a modified FSDM based FPBNN increases from 1 to 3000, its square error \(\hat{e}(k)\) decreases sharply from \(0.82 \times 10^{-3}\) to a zero minimum. As the number of iterations of a modified FSDM based FPBNN increases from one to ten, its square error \(\hat{e}(k)\) decreases sharply from \(3.4 \times 10^{-3}\) to \(0.35 \times 10^{-3}\). When the number of iterations of a modified FSDM based FPBNN increases from 1 to 3000, its square error \(\hat{e}(k)\) decreases sharply from \(0.82 \times 10^{-3}\) to a zero minimum. As the number of iterations of a modified FSDM based FPBNN increases, this minimum remains at zero. Thus, a modified FSDM based FPBNN converges to a global optimal minimum point. From (1), (27), (28), and (79), we observe that when \(0 < v < 1\), the square error \(\hat{e}(k)\) of a modified FSDM based FPBNN may not always decrease; rather, it can increase at certain points, which helps a modified FSDM based FPBNN to bypass the first-order local extreme points of \(\hat{e}(k)\). Figure 7(d) indicates that when the number of iterations of a modified FSDM based FPBNN increases from one to ten, its fractional-order \(v\) decreases sharply from 2.30 to 0.51. When the number of iterations of a modified FSDM based FPBNN increases from 11 to 3000, its fractional-order \(v\) increases gradually from 0.51 to 0.80. Further, when the number of iterations of a modified FSDM based FPBNN increases from 3001 to 4000, its fractional-order \(v\) decreases sharply from 0.80 to zero. From (79), we can observe that the fractional-order \(v\) of a modified FSDM based FPBNN varies with its average error \(\bar{e}(k)\) changing during the entire iterative process. Thus, the fractional-order \(v\) of a modified FSDM based FPBNN in its entire iterative search process is not constant, which demonstrates the fractional-order multi-scale global optimization of a modified FSDM based FPBNN.

Example 3: To further verify the fractional-order multi-scale global optimization of a modified FSDM based FPBNN, we select another initial condition in the sharply sloping zone of the square error \(\hat{e}(k)\) from a random sample of cases. We set the same parameters, the rate of convergence \(\mu = 5.50\), the number of iterations = 6000, for both a modified FSDM based FPBNN and a first-order BPNN in this simulation experiment; the initial condition is at the point \(w_{1j} = 81\) and \(w_{2j} = 9\). From (27)–(30), the iterative search process of a modified FSDM based FPBNN and a first-order BPNN can be represented as displayed in Fig. 8.

![Fig. 8. Comparison of iterative search process of a modified FSDM based FPBNN and a first-order BPNN: (a) Convergence trajectories; (b) Convergence patterns of squared error of \(k^i\) iteration; (c) Fractional-order \(v\) of FPBNN.](image)

In Fig. 8(a), the convergence trajectory of a first-order BPNN illustrates the manner in which it can converge to a local extreme point \((w_{1j} = 0.88\) and \(w_{2j} = 38.60\)). However, the convergence trajectory of a modified FSDM based FPBNN illustrates the manner in which it can converge to a global optimal minimum point \((w_{1j} = 10\) and \(w_{2j} = 1\)). Figure 8(b) indicates that when the number of iterations of a first-order BPNN is approximately equal to 2000, the square error \(\hat{e}(k)\) of a first-order BPNN reduces to a nonzero minimum, which is approximately equal to \(0.25 \times 10^{-1}\). Thus, a first-order BPNN can be trapped in a local extreme point. However, when the number of iterations of a modified FSDM based FPBNN is approximately equal to 2300, the square error \(\hat{e}(k)\) of a modified FSDM based FPBNN reduces to a zero minimum. Thus, a modified FSDM based FPBNN converges to a global optimal minimum point. Figure 8(b) also indicates that when \(0 < v < 1\), the square error \(\hat{e}(k)\) of a modified FSDM based FPBNN may not always decrease; rather, it can increase at certain points, which helps a modified FSDM based FPBNN to bypass the first-order local extreme points of \(\hat{e}(k)\). Figure 8(c) indicates that the fractional-order \(v\) of a modified FSDM based FPBNN in its entire iterative search process is not constant; rather, it varies according to (79), which demonstrates the fractional-order multi-scale global optimization of a modified FSDM based FPBNN.

Example 4: We present an extreme example where the initial condition is directly on a local extreme point of the square error \(\hat{e}(k)\). Assume we have set the same parameters, the rate of convergence \(\mu = 5.50\), the number of iterations = 9000, for both a modified FSDM based FPBNN and a first-order BPNN in this simulation experiment, and the initial condition is directly at the local extreme point \((w_{1j} = 0.88\) and \(w_{2j} = 38.60\)) of \(\hat{e}(k)\). Thus, from (27)–(30), the iterative search process of a modified
FSDM based FBPNN and a first-order BPNN can be represented as displayed in Fig. 9.

Figures 9(a) and 9(e) indicate that if the initial condition is directly at a local extreme point of \( \hat{F}(x) \), the convergence trajectory of a first-order BPNN is trapped at this local extreme point and the square error \( e \) indicates that even if the initial condition is directly at a local extreme point of \( \hat{F}(k) \), the convergence trajectory of a modified FSDM based FBPNN can converge to a global optimal minimum point \( (w_{11} = 10, w_{12} = 1) \). Figure 9(f) indicates that the number of iterations of a modified FSDM based FBPNN increases from one to 3500, its \( \hat{F}(x) \) does not decrease; rather it increases gradually from \( 0.25 \times 10^{-1} \) to \( 7.4 \times 10^{-4} \). Further, when the number of iterations of a modified FSDM based FBPNN increases from 3501 to 5200, the \( \hat{F}(x) \) decreases sharply from \( 7.4 \times 10^{-4} \) to a zero minimum. As the number of iterations of a modified FSDM based FBPNN increases, this zero minimum remains at zero. Thus, a modified FSDM based FBPNN converges to a global optimal minimum point. Figure 9(g) indicates that the fractional-order \( v \) of a modified FSDM based FBPNN in its entire iterative search process is not constant; rather, it varies according to (79), which demonstrates the fractional-order multi-scale global optimization of a modified FSDM based FBPNN.

C. Comparative performance of modified FSDM based FBPNN with real data

In this subsection, to further verify the fractional-order multi-scale global optimization, we implement two comparative performances of a modified FSDM based FBPNN with real data.

Example 5: In this example, the data presented in Table 1 displays the output as a transfer function of the input of a nonlinear signal-processing filter.

We choose a modified FSDM based FBPNN and a Levenberg-Marquardt algorithm-based first-order BPNN [17] to provide two nonlinear least-square approximations to the data set in Table 1. We illustrate the characteristics with a \( 1 - w^1 \) modified FSDM based FBPNN and a same structural first-order BPNN. The number of neurons in the first layer (hidden layer) of a modified FSDM based FBPNN and a first-order BPNN are both equal to \( 151 \). The activation functions for the first layer and second layer are tan-sigmoid \( f'(x) = \frac{1}{(1 + e^{-x})} \) and pure linear function \( f'(x) = x \), respectively. \( w_{11} = [w_{11}, 20.8597, 21.2543, -21.0232, -21.3975, 21.0826, -21.0743, 21.0052, 21.0272, 20.9446, -21.1307, 21.2419, 20.9357, 21.0157]^T \) and \( w_{12} = [-0.7629, -0.7168, 1.1592, 0.4330, 0.9470, 0.5903, -1.1983, -0.7002, -0.3756, -1.0144, -0.2451, -1.3834, 0.4546, 0.2460, 0.3230] \) are randomly selected as the weight matrices of the first and second layer of a modified FSDM based FBPNN and a first-order BPNN, respectively. \( b_{11} = [-21.0070, -18.1627, -14.6449, 11.9684, 8.0087, -5.7329, 2.0816, 0.7399, 2.7071, 6.1967, -8.9802, -11.7774, 14.6532, 18.0707, 20.9846]^T \) and \( b_{12} = [-0.4954] \) are also randomly selected as the bias matrices of the first and second layer of a modified FSDM based FBPNN and a first-order BPNN, respectively. For the convenience of illustration, we simultaneously vary only two parameters of a modified FSDM based FBPNN. Figure 10 illustrates the mean squared error of the FBPNN when only two parameters, \( w_{11}, w_{12} \), are adjusted; the other parameters are set to their aforementioned randomly selected values.

Table 1. Input and output of a nonlinear signal-processing filter.

| Layer | \( b_{11} \)       | \( b_{12} \)       |
|-------|-------------------|-------------------|
| Input | [0.0126, 0.0038, 0.0028] | [0.0126, 0.0038, 0.0028] |
| Output| [0.1200, 0.0200, 0.0020] | [0.1200, 0.0200, 0.0020] |
It can be observed that a relative optimal minimum error occurs when \( w_{i_1} = 19.3065 \) and \( w_{i_1'} = -20.4575 \), as indicated by the solid green circle in Fig. 10. We set the same parameters, the rate of convergence \( \mu = 3.50 \), the number of iterations is equal to 3000, for both a modified FSDM based FBPNN and a first-order BPNN in this simulation experiment and the initial conditions were at the points where \( (w_{i_1} = 108 \) and \( w_{i_1'} = 116 \); \( (w_{i_1} = -110 \) and \( w_{i_1'} = -106 \)), and \( (w_{i_1} = -95 \) and \( w_{i_1'} = 100 \)). From (27)–(30), the iterative search process of a modified FSDM based FBPNN and a first-order BPNN can be represented as in Fig. 11.

In Figs. 11(a), (d), and (g), the convergence trajectories illustrate that a first-order BPNN can converge to three local extreme points, \( (w_{i_1} = 86.0196 \) and \( w_{i_1'} = 84.169 \)), \( (w_{i_1} = -83.8969 \) and \( w_{i_1'} = -74.0297 \)), and \( (w_{i_1} = -94.9789 \) and \( w_{i_1'} = 74.0876 \)). However, all convergence trajectories of illustrate that a modified FSDM based FBPNN can converge to a relative global optimal minimum point \( (w_{i_1} = 19.3065 \) and \( w_{i_1'} = -20.4575 \)). Figures 11(b), (e), and (h) display that at first, when the number of iterations increases, the square error \( \hat{F}(k) \) of a first-order BPNN reduces to a nonzero minimum, which is clearly greater than that of a modified FSDM based FBPNN. Secondly, because we vary only two parameters \((w_{i_1} \text{ and } w_{i_1'})\) simultaneously, the maximum adjustment of a modified FSDM based FBPNN converges to the relative global optimal minimum point \( (w_{i_1} = 19.3065 \) and \( w_{i_1'} = -20.4575 \)), where its minimum square error \( \hat{F}(k) (F_{\text{min}} = 0.001115) \) approaches zero, however, not equal to zero. If we vary all the parameters simultaneously, the convergence trajectory of a modified FSDM based FBPNN can converge to a global optimal minimum point, where its minimum square error \( \hat{F}(k) \) is equal to zero. Figures 11(c), (f), and (h), and (79) indicate that the minimum square error \( \hat{F}(k) (F_{\text{min}} = 0.001115) \) approaches zero and the minimum of the fractional-order \( \nu \) of a modified FSDM based FBPNN approaches zero, however, not equal to zero.

In the following extreme example, the initial condition is directly on a local extreme point of the square error \( \hat{F}(k) \). We set the same parameters, the rate of convergence \( \mu = 3.50 \), the number of iterations be equal to 3000 for both a modified FSDM based FBPNN and a first-order BPNN in this simulation experiment; the initial condition is directly at a local extreme point \( (w_{i_1} = -9.00 \) and \( w_{i_1'} = 8.2676 \)) of \( \hat{F}(k) \). Thus, from (27)–(30), the iterative search process of a modified FSDM based FBPNN and a first-order BPNN can be represented as indicated in Fig. 12.
The convergence trajectory of a modified FSDM based FBPNN can converge to a relative global optimal minimum point \( \hat{w}_{1,1} = 19.3065 \) and \( \hat{w}_{1,1} = -20.4575 \).

The responses of a modified FSDM based FBPNN and a first-order BPNN for the convergence parameters of Fig. 11 are displayed in Fig. 13; this consists of the plots of the outputs of a modified FSDM based FBPNN and a first-order BPNN as the inputs are varied over the range of Table 1.

Figure 12 indicates that if the initial condition is directly at a local extreme point of \( \hat{f}(k) \), the convergence trajectory of a first-order BPNN are trapped at this local extreme point and the square error \( \hat{f}(k) \) remains unchanged. Conversely, even if the initial condition is directly at a local extreme point of \( \hat{f}(k) \), the convergence trajectory of a modified FSDM based FBPNN can converge to a relative global optimal minimum point \( \hat{w}_{1,1} = 19.3065 \) and \( \hat{w}_{1,1} = -20.4575 \).

The responses of a modified FSDM based FBPNN and a first-order BPNN for the convergence parameters: (a) Convergence parameters of Fig. 11(a); (b) Convergence parameters of Fig. 11(d); (c) Convergence parameters of Fig. 11(g).

Figure 13 indicates that for the convergence parameters of Fig. 11, the sample data of Table 1 can be fit well by a modified FSDM based FBPNN. The fitting error of a first-order BPNN is clearly greater than a modified FSDM based FBPNN.

V. CONCLUSIONS

The application of fractional calculus to neural networks and cybernetics is an emerging field of study and only a small number of studies have been conducted in this area. The properties of the fractional calculus of a signal are considerably different from those of its integer-order calculus. Fractional calculus has been applied to neural networks and cybernetics primarily owing to its inherent advantages of long-term memory, nonlocality, and weak singularity. Therefore, to improve the optimization performance of the ordinary a first-order BPNNs, it is logical to generalize a first-order BPNN to a modified FSDM based FBPNN by applying a state-of-the-art application of promising mathematical method, fractional calculus. From this inspiration, in this study, a modified FSDM based FBPNN was achieved whose reverse incremental search was in the negative directions of the approximate fractional-order partial derivatives of the square error \( \hat{f}(k) \). The more efficient optimal searching capability of the fractional-order multi-scale global optimization of a modified FSDM based FBPNN to determine the global optimal solution is the major advantage being superior to a classic first-order BPNN.

From the aforementioned discussion, we can also observe that there are other problems that must be further studied. For example, the aforementioned imperfect adaptive kernel function of the fractional-order \( \nu \) at the \( k^\text{th} \) iteration of a modified FSDM based FBPNN, \( \nu(k) \), is not sufficient for an arbitrary quadratic energy norm \( \hat{f}(k) \). Further, the computational complexity of a modified FSDM based FBPNN increases with the introduction of fractional calculus. With the same number of neurons, the computational complexity of a modified FSDM based FBPNN is 4 times greater than that of a classic first-order BPNN in every iterative computation. Therefore, it is evident that other topics such as how to construct an efficient appropriate correlation function of \( \mu(k) \), how to construct a more efficient adaptive kernel function of the fractional-order \( \nu(k) \), and how to reduce the computational effort of a modified FSDM based FBPNN must be studied further. These topics will be discussed in our future work.

REFERENCES

[1] R. Battiti, “First and Second Order Methods for Learning: Between Steepest Descent and Newton’s Method,” Neural Computation, vol. 4, no. 2, pp. 141-166, 1992.
[2] P. J. Werbos, Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences. Ph.D. Thesis, Harvard University, USA, 1974.
[3] D. E. Rumelhart, G. E. Hinton and R. J. Williams, “Learning Representations by Back-Propagating Errors,” Nature, vol. 323, no. 9, pp. 533-536, 1986.
[4] D. B. Parker, “Learning-Logic: Casting the Cortex of the Human Brain in Silicon,” Center for Computational Research in Economics and Management Science, MIT, USA, Tech. Rep. Technical Report TR-47, 1985.
[5] Y. L. Cun, “Une Procedure d’Apprentissage pour Reseau a Seuil Asymmetrique,” Proceedings of Cognitiva 85: A la Frontiere de l’Intelligence Artificielle des Sciences de la Connaissance des Neurosciences, vol. 85, pp. 599-604, 1985.
[6] D. E. Rumelhart, J. L. McClelland and the PDP Research Group, Parallel Distributed Processing: Explorations in the Microstructure of Cognition, vol. 1, Cambridge: MIT Press, 1986.
