

Fractional-order global optimal backpropagation machine trained by an improved fractional-order steepest descent method

Yi-fei PU (蒲亦非) , Jian WANG (王健)

puyifei@scu.edu.cn, wangjiannl@upc.edu.cn

Yi-fei PU, Jian WANG, 2020. Fractional-order global optimal backpropagation machine trained by an improved fractional-order steepest descent method. *Frontiers of Information Technology & Electronic Engineering*, 21(6):809-833. <https://doi.org/10.1631/FITEE.1900593>



Contents

1

Background

2

Methods

3

Experiments

4

Conclusions



四川大学
SICHUAN UNIVERSITY

计算机学院



PART 1

Background

Front Inform Technol Electron Eng



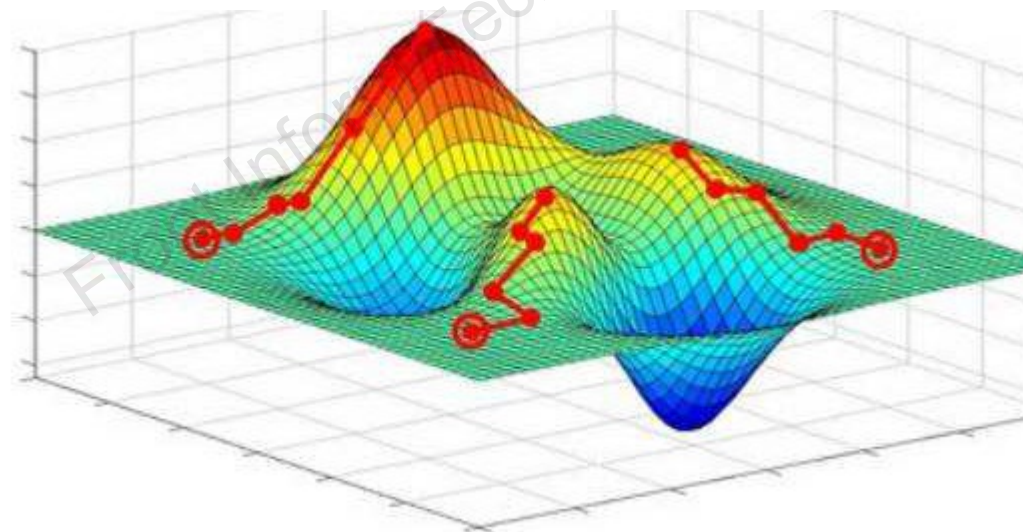
Problem

Classical first-order steepest descent method (SDM)

Simple

Valid in convex functions

Invalid in non-convex functions





Problem

Classical first-order steepest descent method (SDM) based BPNNs

The classical first-order SDM-based BPNNs are easily trapped into **a local optimal solution**. Their optimization performances must be improved.

Front Inform Technol Electron Eng



四川大学
SICHUAN UNIVERSITY

计算机学院



海纳百川，有容乃大

以人为本，崇尚学术

PART 2

Methods

Front Inform Technol Electron Eng



Preparation 1

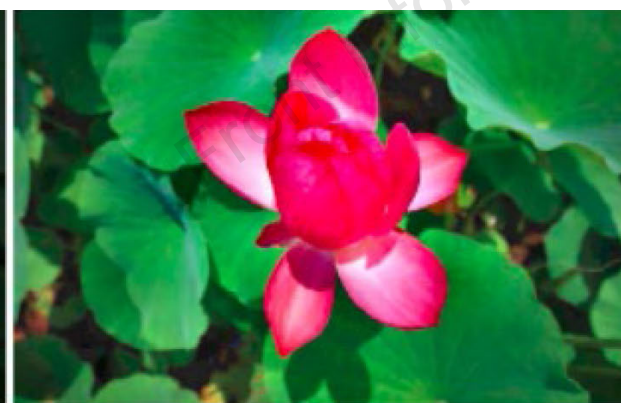
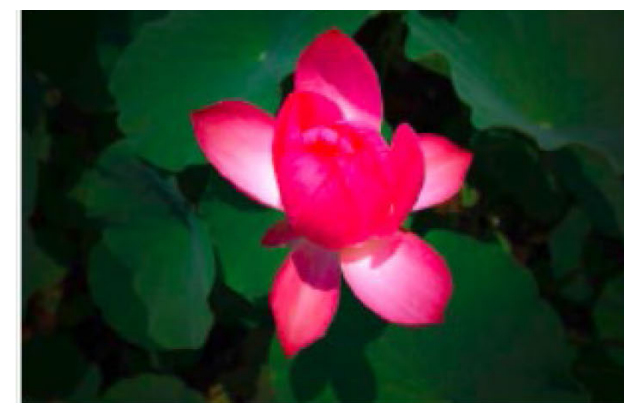
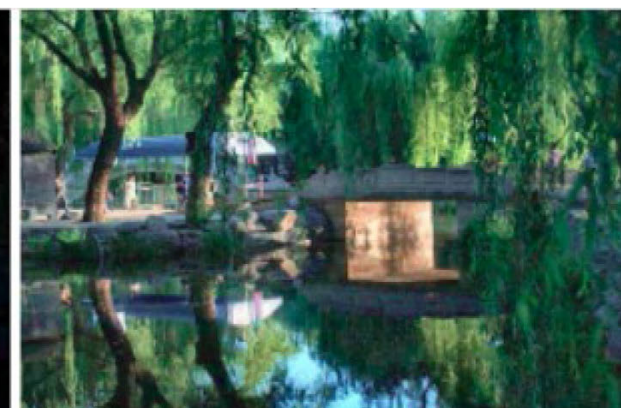
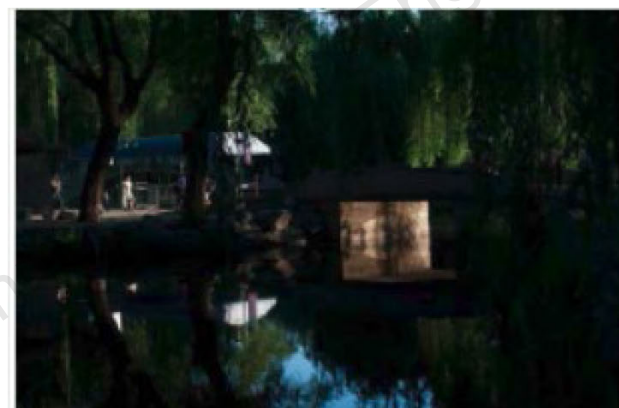
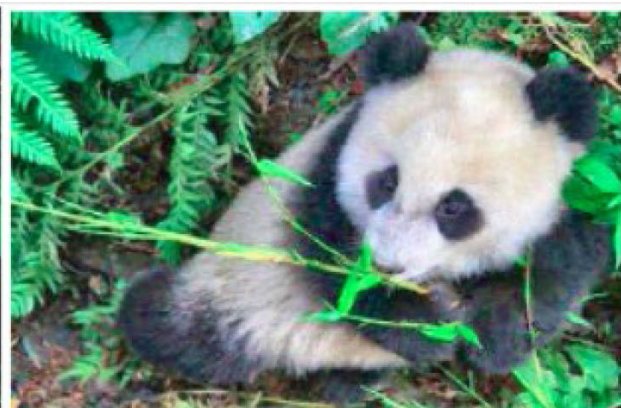
Fractional calculus

Fractional calculus has become an important novel branch in mathematical analyses. Recently, promising results and ideas have demonstrated that fractional calculus can be an interesting and promising tool in many scientific fields such as diffusion processes, viscoelasticity theory, fractal dynamics, fractional control, fractor (fractional-order resistor), fracmemristor (fractional-order memristor), image processing, and neural networks.



Preparation 1

Fractional calculus



Preparation 1

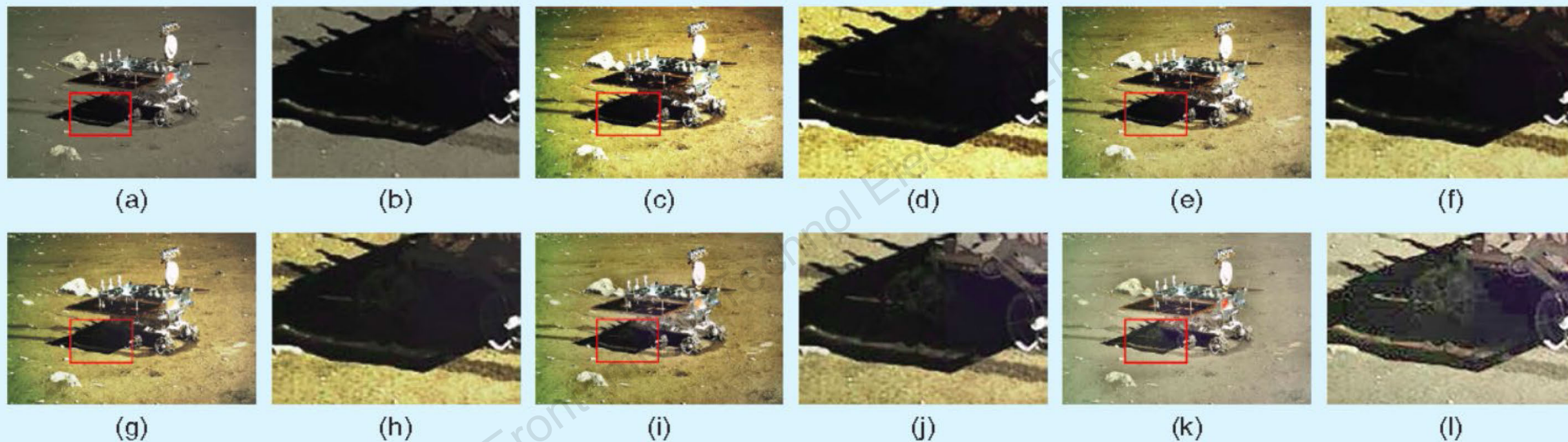


FIG 1 Comparative contrast enhancement experiments of photo of Yutu lunar rover: (a) original image; (b) enlarged detail of central 1/16 of (a); (c) CLAHE of (a); (d) enlarged detail of central 1/16 of (c); (e) SSR of (a); (f) enlarged detail of central 1/16 of (e); (g) MSR of (a); (h) enlarged detail of central 1/16 of (g); (i) VR of (a); (j) enlarged detail of central 1/16 of (i); (k) FR of (a); (l) enlarged detail of central 1/16 of (k).



Preparation 1

Fractional calculus

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.

The application of the improved fractional-order steepest descent method (FSDM) in the training of BPNNs has potential. To **improve the optimization performance of classical first-order BPNNs**, in this work, we study whether it is possible to apply the improved FSDM to generalize classical first-order BPNNs to fractional-order BPNNs (FBPNNs).



Preparation 1

Fractional calculus

The basic characteristic 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 classical integer-order calculus.

Front Inform Technol Electron Eng



Preparation 1

Fractional calculus

Grünwald-Letnikov

Riemann-Liouville

Caputo

$${}^{G-L}D_x^v f(x) = \lim_{N \rightarrow \infty} \left\{ \frac{\left((x-a)/N \right)^{-v}}{\Gamma(-v)} \sum_{k=0}^{N-1} \frac{\Gamma(k-v)}{\Gamma(k+1)} f\left(x - k \left(\frac{x-a}{N} \right) \right) \right\}$$

f is a differintegrable function

$[a, x]$ is the duration of f

N is the number of partitions of the duration

v is an arbitrary real number (fractional-order)

Γ is the Gamma function



Preparation 1

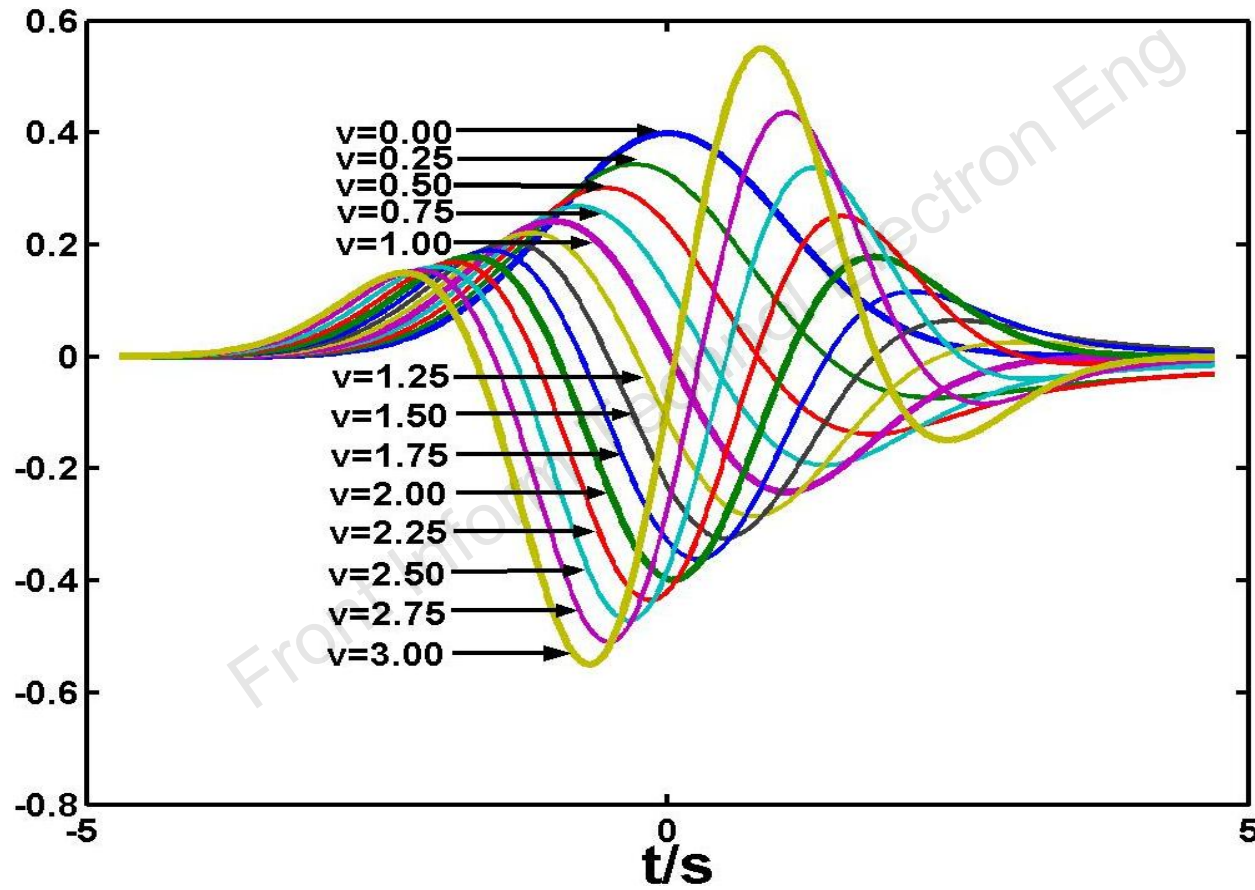
“ a ” is the minimum value of the domain. Therefore, the fractional derivatives of a real-valued function are real numbers, not complex numbers.

$${}^{G-L}D_x^{\nu} f(x) = \lim_{N \rightarrow \infty} \left\{ \frac{\left(\frac{(x-a)}{N} \right)^{-\nu}}{\Gamma(-\nu)} \sum_{k=0}^{N-1} \frac{\Gamma(k-\nu)}{\Gamma(k+1)} f\left(x - k \left(\frac{x-a}{N} \right)\right) \right\}$$

Fractional calculus has the potential of playing an important role in the applications of neural networks and cybernetics because of its inherent strengths such as long-term memory, non-locality, and weak singularity.



Preparation 1





Preparation 1

The fractional differential, except based on the **Caputo** definition, of a Heaviside function, is nonzero, whereas its integer-order differential must be zero.

Thus, the properties of the fractional-order steepest descent method (FSDM) are also different from those of the traditional first-order steepest descent method.

For example, FSDM can determine the fractional-order extreme points of the energy norm, which do not overlap the traditional first-order stationary points. It is known that the classical first-order BPNNs has a tendency to be trapped into local optimal solutions.



Preparation 1

This introduced fractional-order branch of the family of BPNNs trained by the improved FSDM differs from the majority of the previous classical first-order BPNNs. The higher optimal search ability of an FBPNN to **determine the global optimal solution** is the major advantage that makes FBPNN superior to a classical first-order BPNN.

Front Inform Technol Electron Eng



Preparation 2

Fractional calculus

Grünwald-Letnikov numerical implementation

Fractional-order operator/mask

$$\begin{aligned} \frac{d^\nu s(t)}{dt^\nu} \approx & s(t) + (-\nu)s(t-1) + \frac{(-\nu)(-\nu+1)}{2}s(t-2) \\ & + \frac{(-\nu)(-\nu+1)(-\nu+2)}{6}s(t-3) + \dots + \frac{\Gamma(-\nu+1)}{n!\Gamma(-\nu+n+1)}s(t-n) \end{aligned}$$

$\frac{\nu^2-\nu}{2}$	0	$\frac{\nu^2-\nu}{2}$	0	$\frac{\nu^2-\nu}{2}$
0	-V	-V	-V	0
$\frac{\nu^2-\nu}{2}$	-V	8	-V	$\frac{\nu^2-\nu}{2}$
0	-V	-V	-V	0
$\frac{\nu^2-\nu}{2}$	0	$\frac{\nu^2-\nu}{2}$	0	$\frac{\nu^2-\nu}{2}$

Truncation is necessary (effective memory)

On one hand, it can reduce the amount of calculation; on the other hand, it is convenient for discretization processing, and satisfactory results can be obtained.

Preparation 3

Chain rule of the composite function

$$y = f(g(x))$$

Integer-order derivative

$$\frac{dy}{dx} = \frac{df}{dg} \cdot \frac{dg}{dx}$$

Fractional-order derivative

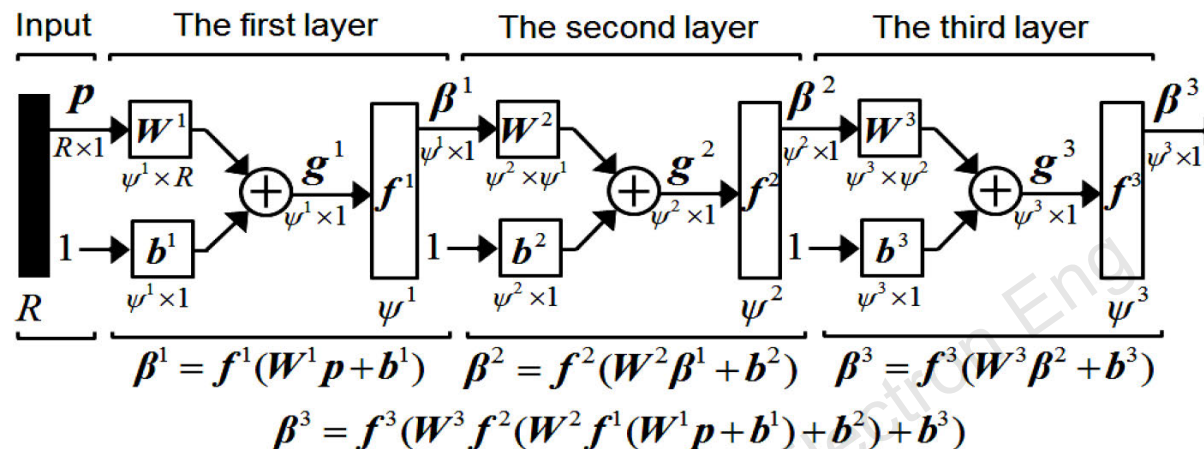
$$\frac{d^{\nu} y}{dx^{\nu}} = \frac{df}{dg} \cdot \frac{d^{\nu} g}{dx^{\nu}}$$



Faà di Bruno's formula

$${}_{x_{\min}}^{GL} D_x^{\nu} f[g(x)] = \frac{(x - x_{\min})^{-\nu}}{\Gamma(1 - \nu)} f[g(x)] + \sum_{j=1}^{\infty} \binom{\nu}{j} \frac{(x - x_{\min})^{j-\nu}}{\Gamma(j - \nu + 1)} j! \sum_{m=1}^j (D_g^m f) \sum_{k=1}^j \frac{1}{P_k!} \left[\frac{D_x^k g}{k!} \right]^{P_k}$$

$$\begin{cases} \sum_{k=1}^j k P_k = j \\ \sum_{k=1}^j P_k = m \end{cases}$$

FBPNN


Model of a BPNN represented by abbreviated symbols denoting its three layers

Mean square error

$$F(\mathbf{x}) = E(\mathbf{e}^T \mathbf{e}) = E[(\mathbf{q} - \boldsymbol{\beta})^T (\mathbf{q} - \boldsymbol{\beta})]$$

Square error at iteration k

$$\hat{F}(k) = [\mathbf{q}(k) - \boldsymbol{\beta}(k)]^T [\mathbf{q}(k) - \boldsymbol{\beta}(k)] = \mathbf{e}(k)^T \mathbf{e}(k), \quad \hat{F}(k) = \sum_{j=1}^{\psi^M} (q_j - \beta_j)^2$$

\mathbf{p} Input matrix of samples

\mathbf{q} Ideal output matrix of samples

$\boldsymbol{\beta}$ Practical output matrix of samples

W^1, W^2, \dots, W^3

b^1, b^2, \dots, b^3

f^1, f^2, \dots, f^3

Weight matrix

Bias matrix

Activation function



FBPNN

Partial derivatives of the net input signal

$$g_i^m(w_{i,j}^m, b_i^m) = \sum_{j=1}^{\psi^{m-1}} w_{i,j}^m \beta_j^{m-1} + b_i^m.$$

$$D_{w_{i,j}^m}^1 g_i^m(w_{i,j}^m, b_i^m) = \beta_j^{m-1},$$

$$D_{b_i^m}^1 g_i^m(w_{i,j}^m, b_i^m) = 1.$$

$$D_{w_{i,j}^m}^{k>1} g_i^m(w_{i,j}^m, b_i^m) \equiv 0,$$

$$D_{b_i^m}^{k>1} g_i^m(w_{i,j}^m, b_i^m) \equiv 0.$$

The net input signal of the activation function



Fractional-order partial derivatives of the square error

Faà di Bruno's formula

$$D_{w_{i,j}^m}^\nu \hat{F} [g_i^m(w_{i,j}^m, b_i^m)] = \frac{(w_{i,j}^m - w_{\text{inf}})^{-\nu}}{\Gamma(1-\nu)} \hat{F} + \sum_{n=1}^{\infty} \binom{\nu}{n} \frac{(w_{i,j}^m - w_{\text{inf}})^{n-\nu}}{\Gamma(n-\nu+1)} \left(D_{g_i^m}^n \hat{F} \right) (\beta_j^{m-1})^n,$$

$$D_{b_i^m}^\nu \hat{F} [g_i^m(w_{i,j}^m, b_i^m)] = \frac{(b_i^m - b_{\text{inf}})^{-\nu}}{\Gamma(1-\nu)} \hat{F} + \sum_{n=1}^{\infty} \binom{\nu}{n} \frac{(b_i^m - b_{\text{inf}})^{n-\nu}}{\Gamma(n-\nu+1)} \left(D_{g_i^m}^n \hat{F} \right).$$



FBPNN

Approximate fractional-order partial derivatives of the square error

Truncation

The approximate fractional-order partial derivatives of the square error are suggested to merely take the first four terms of two aforementioned equations into account:

$$\tilde{D}_{w_{i,j}^m}^{\nu} \hat{F} [g_i^m(w_{i,j}^m, b_i^m)] = \frac{(w_{i,j}^m - w_{\text{inf}})^{-\nu}}{\Gamma(1-\nu)} \hat{F} + \sum_{n=1}^3 \binom{\nu}{n} \frac{(w_{i,j}^m - w_{\text{inf}})^{n-\nu}}{\Gamma(n-\nu+1)} \left(D_{g_i^m}^n \hat{F} \right) (\beta_j^{m-1})^n,$$

$$\tilde{D}_{b_i^m}^{\nu} \hat{F} [g_i^m(w_{i,j}^m, b_i^m)] = \frac{(b_i^m - b_{\text{inf}})^{-\nu}}{\Gamma(1-\nu)} \hat{F} + \sum_{n=1}^3 \binom{\nu}{n} \frac{(b_i^m - b_{\text{inf}})^{n-\nu}}{\Gamma(n-\nu+1)} \left(D_{g_i^m}^n \hat{F} \right).$$

Improved fractional-order steepest descent method based FBPNN

Therefore, we introduce an FBPNN trained by an improved FSDM, 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.

$$\tilde{D}_{w_{i,j}^m}^{\nu} \hat{F} [g_i^m(w_{i,j}^m, b_i^m)] = \frac{(w_{i,j}^m - w_{\text{inf}})^{-\nu}}{\Gamma(1-\nu)} \hat{F} + \sum_{n=1}^3 \binom{\nu}{n} \frac{(w_{i,j}^m - w_{\text{inf}})^{n-\nu}}{\Gamma(n-\nu+1)} \left(D_{g_i^m}^n \hat{F} \right) (\beta_j^{m-1})^n,$$

$$\tilde{D}_{b_i^m}^{\nu} \hat{F} [g_i^m(w_{i,j}^m, b_i^m)] = \frac{(b_i^m - b_{\text{inf}})^{-\nu}}{\Gamma(1-\nu)} \hat{F} + \sum_{n=1}^3 \binom{\nu}{n} \frac{(b_i^m - b_{\text{inf}})^{n-\nu}}{\Gamma(n-\nu+1)} \left(D_{g_i^m}^n \hat{F} \right).$$

$$w_{i,j}^m(k+1) = w_{i,j}^m(k) - \mu \left[\tilde{D}_{w_{i,j}^m}^{\nu} \hat{F}(k) \right],$$

$$b_i^m(k+1) = b_i^m(k) - \mu \left[\tilde{D}_{b_i^m}^{\nu} \hat{F}(k) \right],$$



FBPNN

Fractional-order global optimal convergence of the improved FSDM based FBPNN

The criteria for the convergence of FBPNNs are:

$$\left\{ \begin{aligned} \tilde{D}_{w_{i,j}^m}^{\nu} \hat{F}(k) &= \left\{ \frac{[w_{i,j}^m(k) - w_{\text{inf}}]^{\nu}}{\Gamma(1-\nu)} \hat{F}(k) + \sum_{n=1}^3 \binom{\nu}{n} \frac{[w_{i,j}^m(k) - w_{\text{inf}}]^{n-\nu}}{\Gamma(n-\nu+1)} {}_n\rho_i^m(k) (\beta_j^{m-1})^n \right\} = 0 \\ \tilde{D}_{b_i^m}^{\nu} \hat{F}(k) &= \frac{[b_i^m(k) - b_{\text{inf}}]^{-\nu}}{\Gamma(1-\nu)} \hat{F}(k) + \sum_{n=1}^3 \binom{\nu}{n} \frac{[b_i^m(k) - b_{\text{inf}}]^{n-\nu}}{\Gamma(n-\nu+1)} {}_n\rho_i^m(k) = 0 \end{aligned} \right. .$$



FBPNN

Fractional-order global optimal convergence of improved FSDM based FBPNN

$$\beta^{m+1} = f^{m+1}(g^{m+1}) = f^{m+1}(W^{m+1} \beta^m + b^{m+1}), \quad m=0,1,\dots,M-1,$$

$$g_i^m(w_{i,j}^m, b_i^m) = \sum_{j=1}^{\psi^{m-1}} w_{i,j}^m \beta_j^{m-1} + b_i^m.$$



$$D_{g_j^m}^1 g_i^{m+1} = w_{i,j}^{m+1} \left(D_{g_j^m}^1 \beta_j^m \right) = w_{i,j}^{m+1} \left[D_{g_j^m}^1 f^m(g_j^m) \right] = w_{i,j}^{m+1} \left[D_{g_j^m}^1 f^m \left(\sum_{k=1}^{\psi^{m-1}} w_{j,k}^m \beta_k^{m-1} + b_j^m \right) \right].$$

FBPNN

Fractional-order global optimal convergence of improved FSDM based FBPNN



$${}_n \rho_i^m(k) = \left(D_{g_j^m}^1 g_i^{m+1} \right)^n \left(D_{g_i^{m+1}}^n \hat{F}(k) \right) = \left\{ w_{i,j}^{m+1} \left[D_{g_j^m}^1 f^m \left(\sum_{k=1}^{\psi^{m-1}} w_{j,k}^m \beta_k^{m-1} + b_j^m \right) \right] \right\}^n {}_n \rho_i^{m+1}(k).$$

$${}_1 \rho_i^M(k) = D_{g_i^M}^1 \hat{F}(k) = D_{g_i^M}^1 \left[\sum_{j=1}^{\psi^M} (q_j - \beta_j)^2 \right] = -2(q_i - \beta_i) \left(D_{g_i^M}^1 \beta_i \right),$$

$$D_{g_i^M}^1 \beta_i = D_{g_i^M}^1 \beta_i^M = D_{g_i^M}^1 f^M(g_i^M).$$



FBPNN

Fractional-order global optimal convergence of improved FSDM based FBPNN



$${}_1\rho_i^M(k) = -2(q_i - \beta_i)(D_{g_i^M}^1 \beta_i) = -2(q_i - \beta_i)[D_{g_i^M}^1 f^M(g_i^M)].$$

$${}_2\rho_i^M(k) = D_{g_i^M}^1 [{}_1\rho_i^M(k)] = -2[(q_i - \beta_i)(D_{g_i^M}^2 \beta_i) - (D_{g_i^M}^1 \beta_i)^2],$$

$${}_3\rho_i^M(k) = D_{g_i^M}^1 [{}_2\rho_i^M(k)] = -2[(q_i - \beta_i)(D_{g_i^M}^3 \beta_i) - 3(D_{g_i^M}^1 \beta_i)(D_{g_i^M}^2 \beta_i)].$$



FBPNN

Fractional-order global optimal convergence of improved FSDM based FBPNN



A necessary condition of the criteria for the convergence of FBPNN is:

$$\left\{ \begin{array}{l} \hat{F}(k) = \sum_{j=1}^{\psi^M} (q_j - \beta_j)^2 = 0 \\ {}_1\rho_i^M(k) = -2(q_i - \beta_i)(D_{g_i^M}^1 \beta_i) = 0 \\ {}_2\rho_i^M(k) = -2\left[(q_i - \beta_i)(D_{g_i^M}^2 \beta_i) - (D_{g_i^M}^1 \beta_i)^2\right] = 0 \\ {}_3\rho_i^M(k) = -2\left[(q_i - \beta_i)(D_{g_i^M}^3 \beta_i) - 3(D_{g_i^M}^1 \beta_i)(D_{g_i^M}^2 \beta_i)\right] = 0 \end{array} \right. ,$$

where $\beta_i = \beta_i^M$ and $i=1, 2, \dots, \psi^M$.



FBPNN

Fractional-order global optimal convergence of improved FSDM based FBPNN

A necessary condition of the criteria for the convergence of FBPNN is:



$$\begin{cases} q_i - \beta_i = 0 \\ D_{g_i^M}^1 \beta_i = 0 \end{cases}$$



$$\begin{cases} \hat{F}(k) = \sum_{j=1}^{\psi^M} (q_j - \beta_j)^2 = 0 \\ {}_1\rho_i^M(k) = D_{g_i^M}^1 \hat{F}(k) = -2(q_i - \beta_i)(D_{g_i^M}^1 \beta_i) = 0 \end{cases},$$



FBPNN

Fractional-order global optimal convergence of improved FSDM based FBPNN

A necessary condition of the criteria for the convergence of FBPNN is:



$$\begin{cases} \hat{F}(k) = 0 \\ {}_1\rho_i^m(k) = D_{g_i^m}^1 \hat{F}(k) = 0 \end{cases},$$

where $m = 0, 1, \dots, M - 1$

$$D_{w_{i,j}^m}^1 \hat{F}(k) = {}_1\rho_i^m(k) \beta_j^{m-1} = \left[D_{g_i^m}^1 \hat{F}(k) \right] \beta_j^{m-1}$$

$$\text{and } D_{b_i^m}^1 \hat{F}(k) = {}_1\rho_i^m(k) = D_{g_i^m}^1 \hat{F}(k)$$

Except saddle points, the improved FSDM can **guarantee** that an FBPNN **converges to a fractional-order optimal minimum** $(w_{i,j}^{m*}, b_i^{m*})$, **a global minimum point**, of the square error $\hat{F}(k)$.



FBPNN

Approximate fractional-order partial derivatives of the square error

$$\begin{cases} \hat{F}(k) = \sum_{j=1}^{\psi^M} (q_j - \beta_j)^2 = 0 \\ {}_1\rho_i^M(k) = D_{g_i}^1 \hat{F}(k) = -2(q_i - \beta_i)(D_{g_i}^1 \beta_i) = 0 \end{cases}, \quad \rightarrow \quad \begin{cases} q_i - \beta_i = 0 \\ D_{g_i}^1 \beta_i = 0 \end{cases}$$

$$\begin{cases} \hat{F}(k) = \sum_{j=1}^{\psi^M} (q_j - \beta_j)^2 = 0 \\ {}_1\rho_i^M(k) = -2(q_i - \beta_i)(D_{g_i}^1 \beta_i) = 0 \\ {}_2\rho_i^M(k) = -2\left[(q_i - \beta_i)(D_{g_i}^2 \beta_i) - (D_{g_i}^1 \beta_i)^2\right] = 0 \\ {}_3\rho_i^M(k) = -2\left[(q_i - \beta_i)(D_{g_i}^3 \beta_i) - 3(D_{g_i}^1 \beta_i)(D_{g_i}^2 \beta_i)\right] = 0 \end{cases},$$

FBPNN

Approximate fractional-order partial derivatives of the square error

$$\tilde{D}_{w_{i,j}^m}^{\nu} \hat{F} [g_i^m(w_{i,j}^m, b_i^m)] = \frac{(w_{i,j}^m - w_{\text{inf}})^{-\nu}}{\Gamma(1-\nu)} \hat{F} + \sum_{n=1}^3 \binom{\nu}{n} \frac{(w_{i,j}^m - w_{\text{inf}})^{n-\nu}}{\Gamma(n-\nu+1)} \left(D_{g_i^m}^n \hat{F} \right) (\beta_j^{m-1})^n,$$

$$\tilde{D}_{b_i^m}^{\nu} \hat{F} [g_i^m(w_{i,j}^m, b_i^m)] = \frac{(b_i^m - b_{\text{inf}})^{-\nu}}{\Gamma(1-\nu)} \hat{F} + \sum_{n=1}^3 \binom{\nu}{n} \frac{(b_i^m - b_{\text{inf}})^{n-\nu}}{\Gamma(n-\nu+1)} \left(D_{g_i^m}^n \hat{F} \right).$$



To further simplify the calculation, without loss of generality, in the actual computations of an FBPNN trained by an improved FSDM, we can set $n=1$.

$$\tilde{D}_{w_{i,j}^m}^{\nu} \hat{F} [g_i^m(w_{i,j}^m, b_i^m)] = \frac{(w_{i,j}^m - w_{\text{inf}})^{-\nu}}{\Gamma(1-\nu)} \hat{F} + \binom{\nu}{1} \frac{(w_{i,j}^m - w_{\text{inf}})^{1-\nu}}{\Gamma(2-\nu)} \left(D_{g_i^m}^1 \hat{F} \right) (\beta_j^{m-1}),$$

$$\tilde{D}_{b_i^m}^{\nu} \hat{F} [g_i^m(w_{i,j}^m, b_i^m)] = \frac{(b_i^m - b_{\text{inf}})^{-\nu}}{\Gamma(1-\nu)} \hat{F} + \binom{\nu}{1} \frac{(b_i^m - b_{\text{inf}})^{1-\nu}}{\Gamma(n-\nu+1)} \left(D_{g_i^m}^1 \hat{F} \right).$$



Comparison 1

BPNN

$$D_{w_{i,j}^m}^1 \hat{F}(k) = \left[D_{g_i^m}^1 \hat{F}(k) \right] \beta_j^{m-1}$$

$$D_{b_i^m}^1 \hat{F}(k) = D_{g_i^m}^1 \hat{F}(k)$$

$$w_{i,j}^m(k+1) = w_{i,j}^m(k) - \mu \left[D_{w_{i,j}^m}^1 \hat{F}(k) \right],$$

$$b_i^m(k+1) = b_i^m(k) - \mu \left[D_{b_i^m}^1 \hat{F}(k) \right],$$

FBPNN

$$\tilde{D}_{w_{i,j}^m}^v \hat{F}(k) = \frac{(w_{i,j}^m - w_{\text{inf}})^{-v}}{\Gamma(1-v)} \hat{F} + \binom{v}{1} \frac{(w_{i,j}^m - w_{\text{inf}})^{1-v}}{\Gamma(2-v)} \left(D_{g_i^m}^1 \hat{F} \right) (\beta_j^{m-1}),$$

$$\tilde{D}_{b_i^m}^v \hat{F}(k) = \frac{(b_i^m - b_{\text{inf}})^{-v}}{\Gamma(1-v)} \hat{F} + \binom{v}{1} \frac{(b_i^m - b_{\text{inf}})^{1-v}}{\Gamma(n-v+1)} \left(D_{g_i^m}^1 \hat{F} \right).$$

$$w_{i,j}^m(k+1) = w_{i,j}^m(k) - \mu \left[\tilde{D}_{w_{i,j}^m}^v \hat{F}(k) \right],$$

$$b_i^m(k+1) = b_i^m(k) - \mu \left[\tilde{D}_{b_i^m}^v \hat{F}(k) \right],$$

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 an FBPNN is **similar** to that of a traditional BPNN.

Simple

Small computational burden

Small storage space



Comparison 1

BPNN

$$D_{w_{i,j}}^1 \hat{F}(k) = \left[D_{g_i}^1 \hat{F}(k) \right] \beta_j^{m-1}$$

$$D_{b_i}^1 \hat{F}(k) = D_{g_i}^1 \hat{F}(k)$$

$$w_{i,j}^m(k+1) = w_{i,j}^m(k) - \mu \left[D_{w_{i,j}}^1 \hat{F}(k) \right],$$

$$b_i^m(k+1) = b_i^m(k) - \mu \left[D_{b_i}^1 \hat{F}(k) \right],$$

Stop rule

FBPNN

$$\tilde{D}_{w_{i,j}}^v \hat{F}(k) = \frac{(w_{i,j}^m - w_{\text{inf}})^{-v}}{\Gamma(1-v)} \hat{F} + \binom{v}{1} \frac{(w_{i,j}^m - w_{\text{inf}})^{1-v}}{\Gamma(2-v)} \left(D_{g_i}^1 \hat{F} \right) (\beta_j^{m-1}),$$

$$\tilde{D}_{b_i}^v \hat{F}(k) = \frac{(b_i^m - b_{\text{inf}})^{-v}}{\Gamma(1-v)} \hat{F} + \binom{v}{1} \frac{(b_i^m - b_{\text{inf}})^{1-v}}{\Gamma(n-v+1)} \left(D_{g_i}^1 \hat{F} \right).$$

$$w_{i,j}^m(k+1) = w_{i,j}^m(k) - \mu \left[\tilde{D}_{w_{i,j}}^v \hat{F}(k) \right],$$

$$b_i^m(k+1) = b_i^m(k) - \mu \left[\tilde{D}_{b_i}^v \hat{F}(k) \right],$$

The number of iterations is greater than a threshold.

The loss function is less than an acceptable threshold.

The first-order derivative is equal to 0.

The number of iterations is greater than a threshold.

The loss function is less than an acceptable threshold.

The first-order derivative is equal to 0.

+

The square error function is equal to 0.



FBPNN

Fractional-order global optimal convergence of the improved FSDM based FBPNN

Lemma 1 If the number of neurons and number of hidden layers of an FBPNN trained by an improved FSDM are such that at least one fractional-order minimum point of the square error exists, even if $\tilde{D}_{w_{i,j}^m}^v \hat{F}(k) = 0$ and $\tilde{D}_{b_i^m}^v \hat{F}(k) = 0$, but $\hat{F}(k) \neq 0$, the iterative search processes are constrained to keep going on. Then the improved FSDM can **guarantee** that an FBPNN **converges to a fractional-order optimal minimum point** $(w_{i,j}^{m*}, b_i^{m*})$, **a global minimum point.**



Comparison 2

Lemma 1 indicates that first a classical first-order BPNN considers merely the **local characteristics** of the square error F . Thus, a classical first-order BPNN is likely to converge to a local extreme point of its square error.

Assume that the iterative search process of a classical first-order BPNN converges to a local extreme point. On this **local extreme point**, one can obtain

$$\begin{cases} \hat{F}(k) \neq 0 \\ {}_1\rho_i^m(k) = D_{g_i^m}^1 \hat{F}(k) = 0 \end{cases}$$



Comparison 2

$$\text{BPNN} \begin{cases} \hat{F}(k) \neq 0 \\ {}_1\rho_i^m(k) = D_{g_i^m}^1 \hat{F}(k) = 0 \end{cases}$$

$$\text{FBPNN} \begin{cases} \hat{F}(k) = 0 \\ {}_1\rho_i^m(k) = D_{g_i^m}^1 \hat{F}(k) = 0 \end{cases}$$

However, our method considers the **nonlocal characteristics and weak singularity** of the square error, F , of an FBPNN. Thus, an FBPNN trained by an improved FSDM can guarantee the convergence to a global minimum point of its square error.

The higher optimal search ability of an FBPNN to determine the **global optimal solution** is **the major advantage** that makes FBPNN superior to a classical first-order BPNN.

Valid in convex functions

Valid in non-convex functions



Comparison 3

In general, the fractional-order extreme value of a normalized quadratic energy norm determined by the fractional-order partial derivatives is not equal to its integer-order one.

How to solve this problem ?

If the global minimum value of the square error is equal to zero, the fractional-order optimal minimum point, determined by the approximate fractional-order partial derivatives, is identical to the global minimum value of the square error.



Comparison 3

The universal approximation theorem states that the ideal value of the error function equals 0.

However, in practice, this value is usually greater than 0.

So, when should we stop?

The engineering problem, **on one hand**, is based on the operator's experience, and **on the other hand**, can be determined by the needs of the customer service.

FBPNN

Can find **arbitrarily** the position corresponding to the given error.

BPNN

Can NOT find.



四川大学
SICHUAN UNIVERSITY

计算机学院



海纳百川，有容乃大

以人为本，崇尚学术

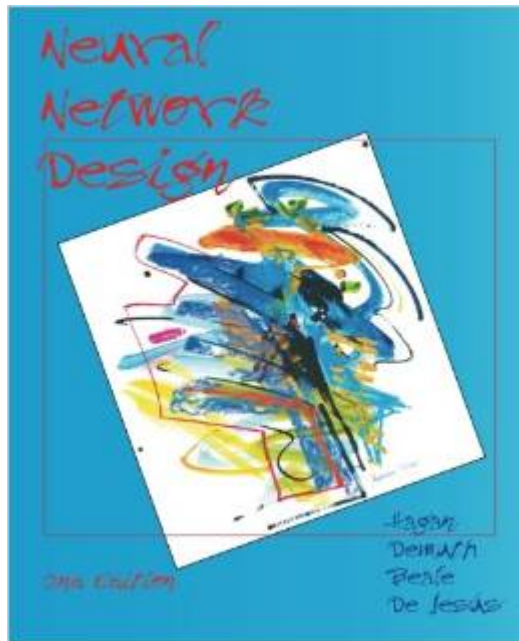
PART 3

Experiments

Front Inform Technol Electron Eng



Example 1



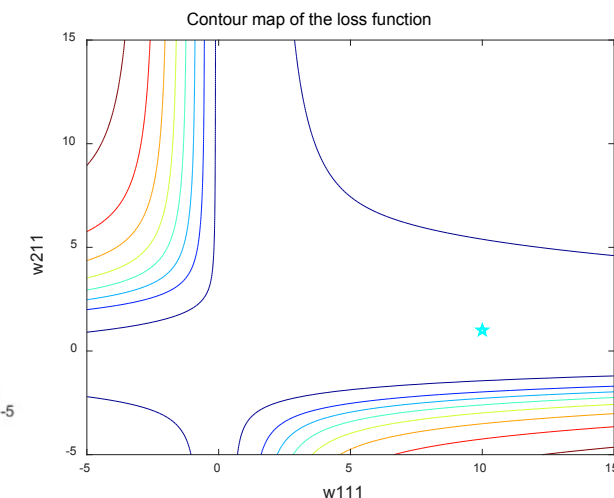
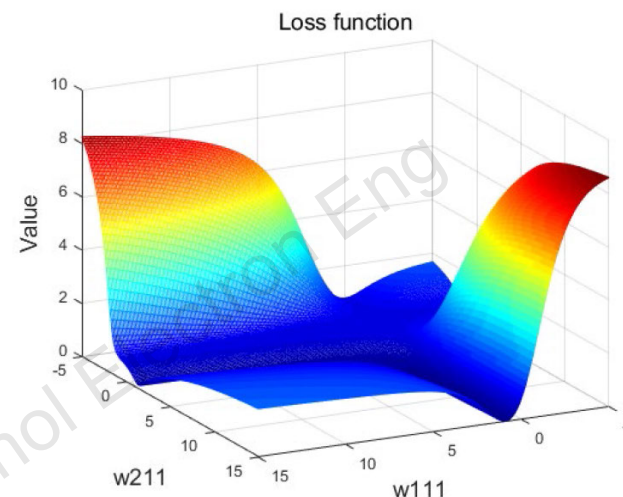
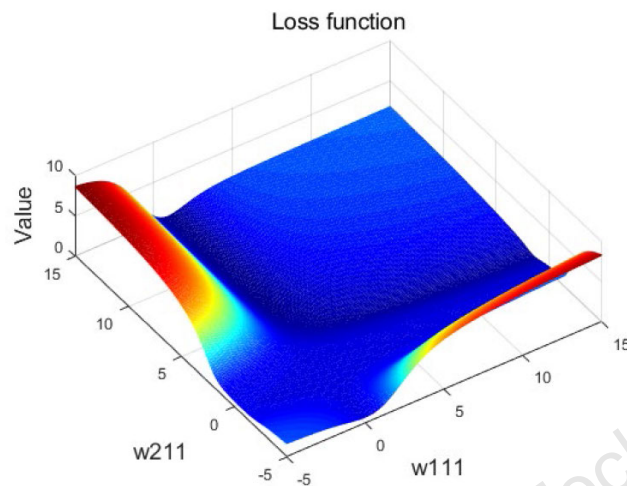
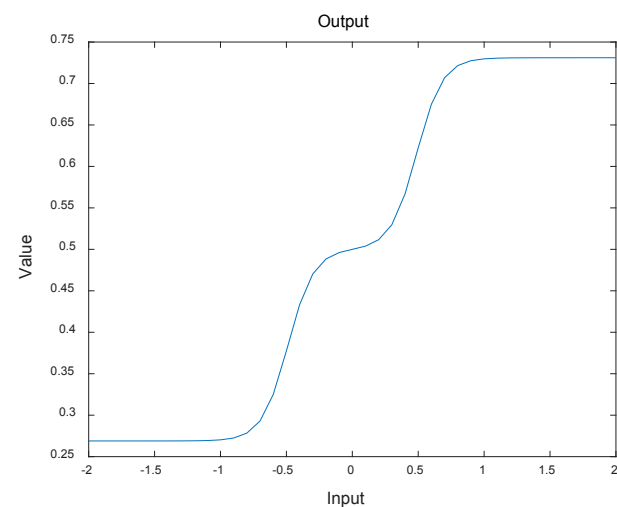
Neural Network Design (2nd Edition)

Martin T. Hagan, Howard B. Demuth, Mark H. Beale, Orlando De Jesus

ISBN-10: 0-9717321-1-6

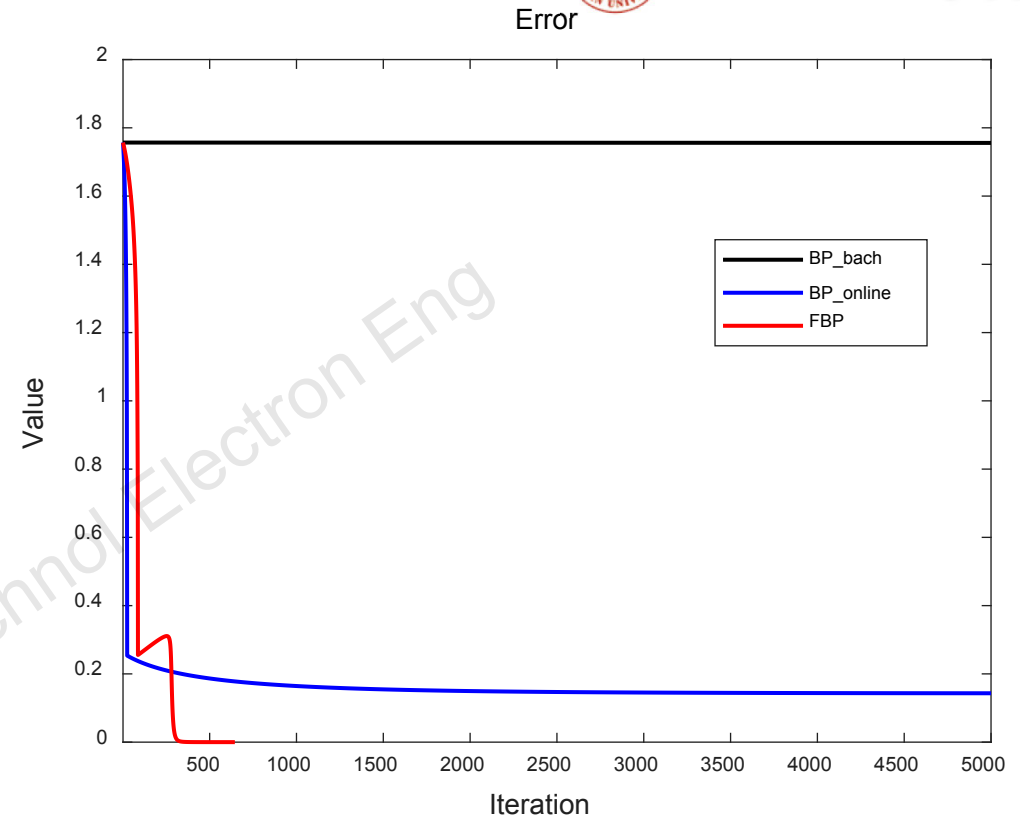
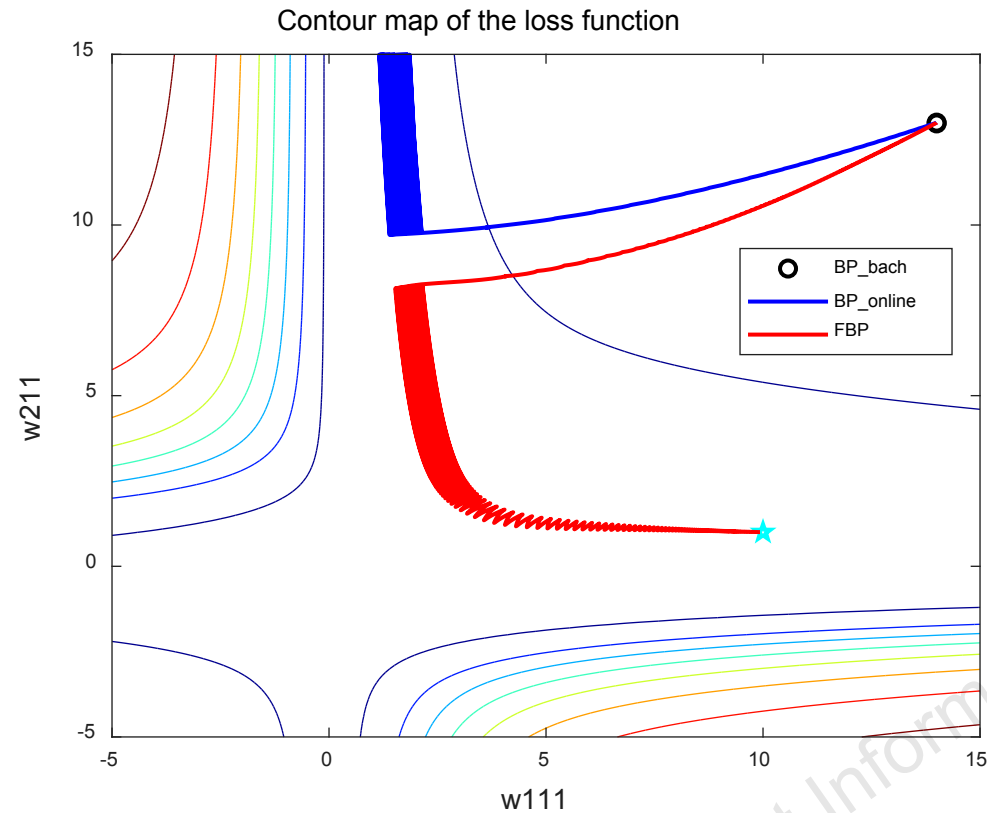
ISBN-13: 978-0-9717321-1-7

<http://hagan.okstate.edu/nnd.html>



What do you think of this terrain?

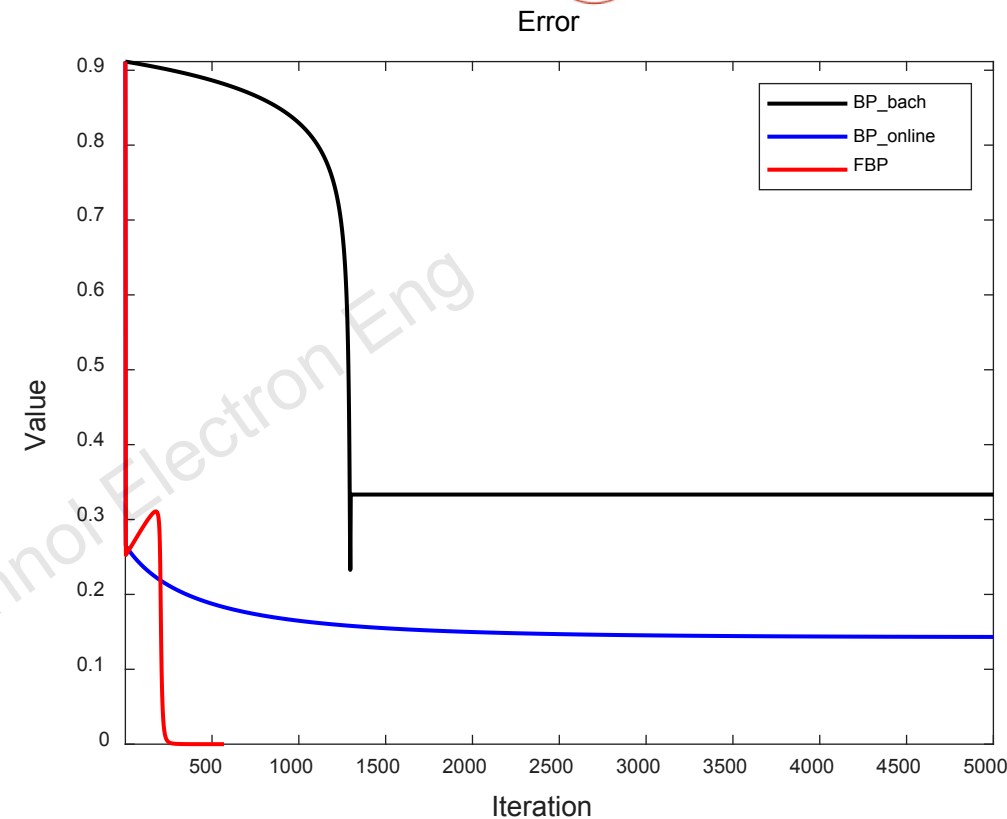
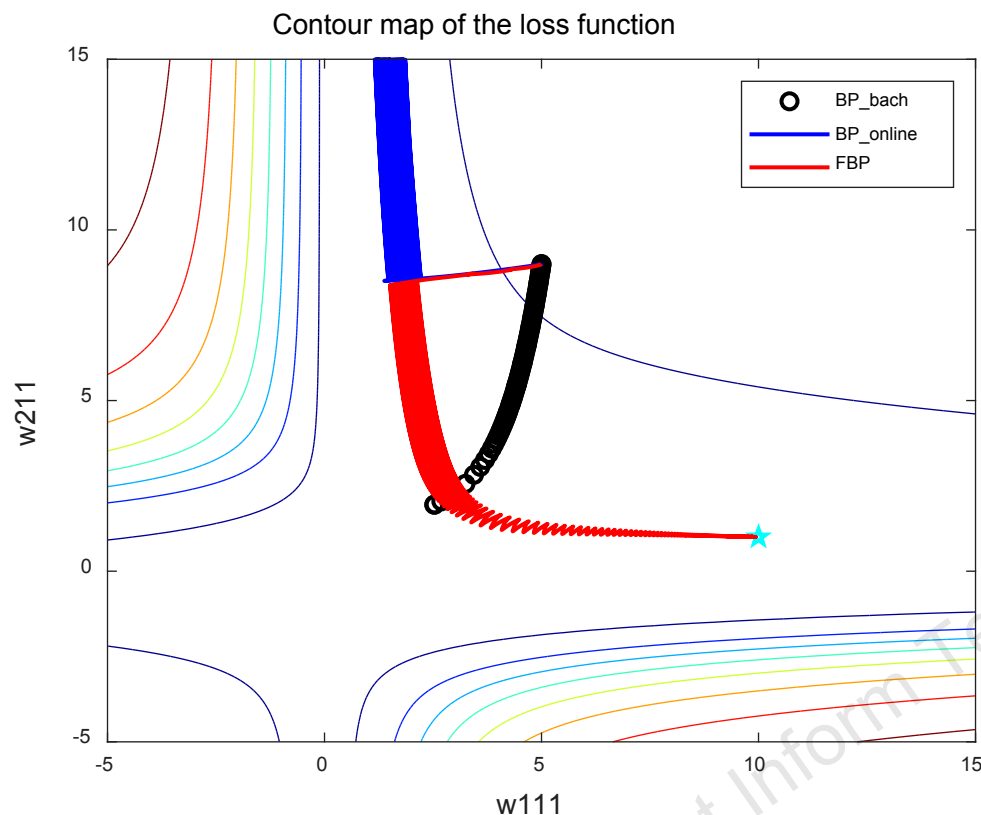
Is it not simple?



The surrounding terrain is very flat, so the integer-order derivative is close to zero.

Gradient does NOT disappear

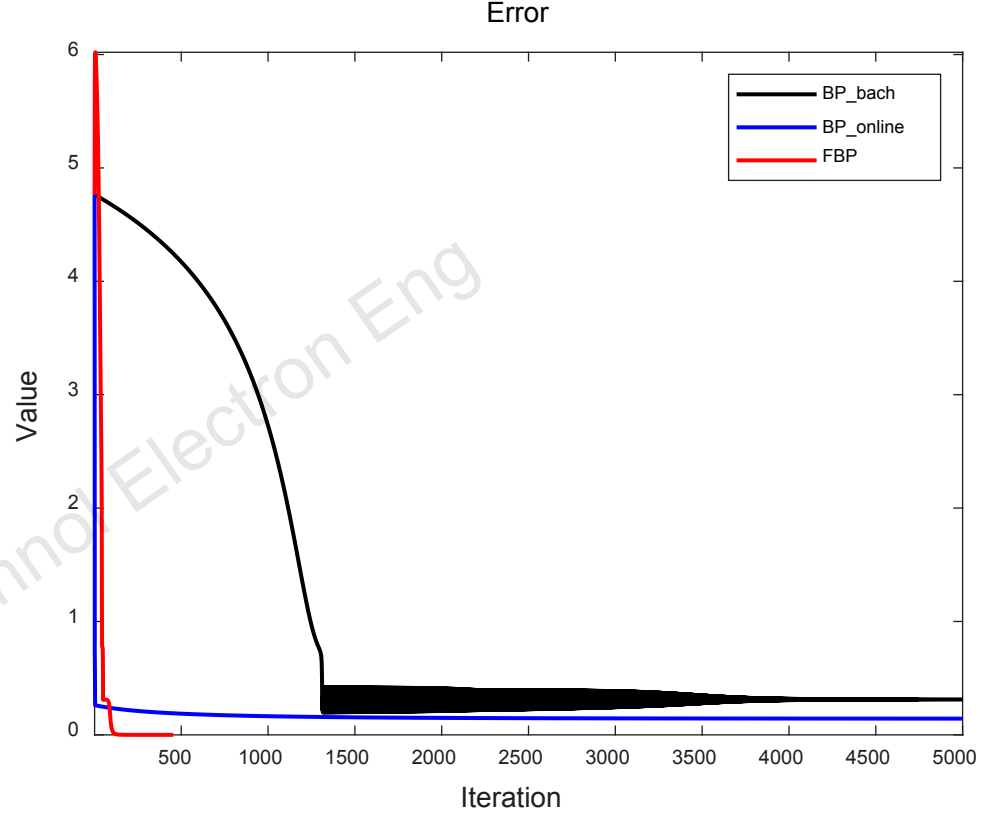
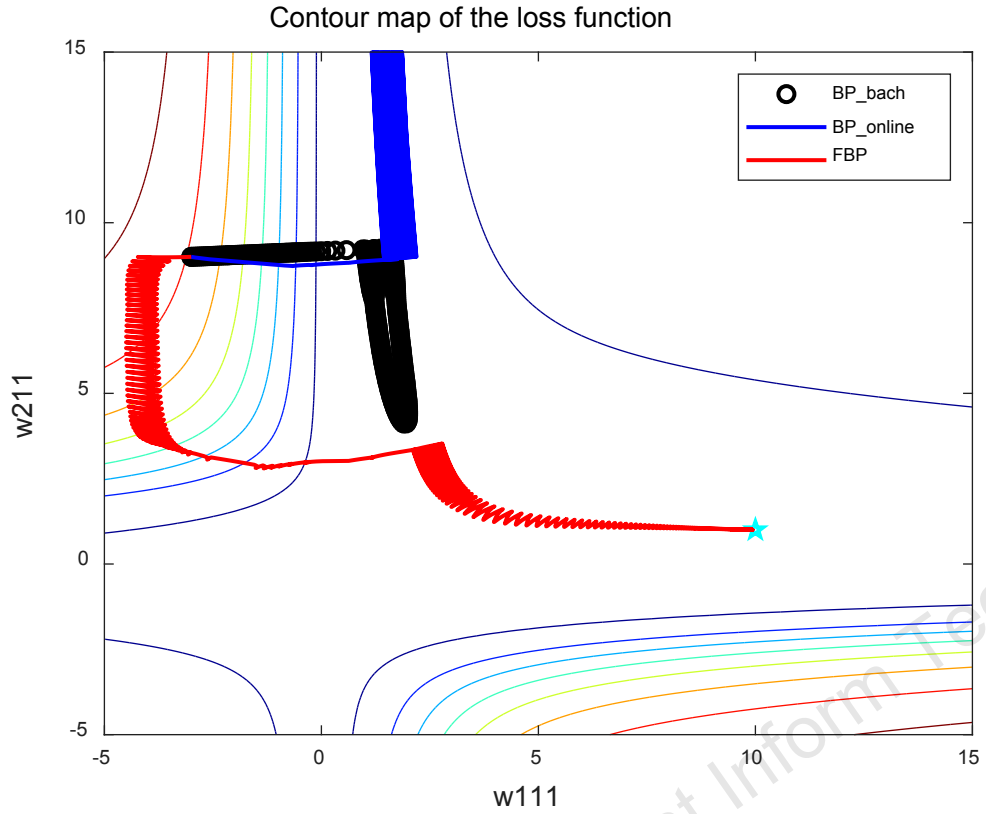
FBPNN can converge to the global optimal point, and its convergence speed is **very fast**.



The middle terrain also has many special points with an integer-order derivative close to zero.

FBPNN can converge to the global optimal point, and its convergence speed is **very fast**.

Stable running results



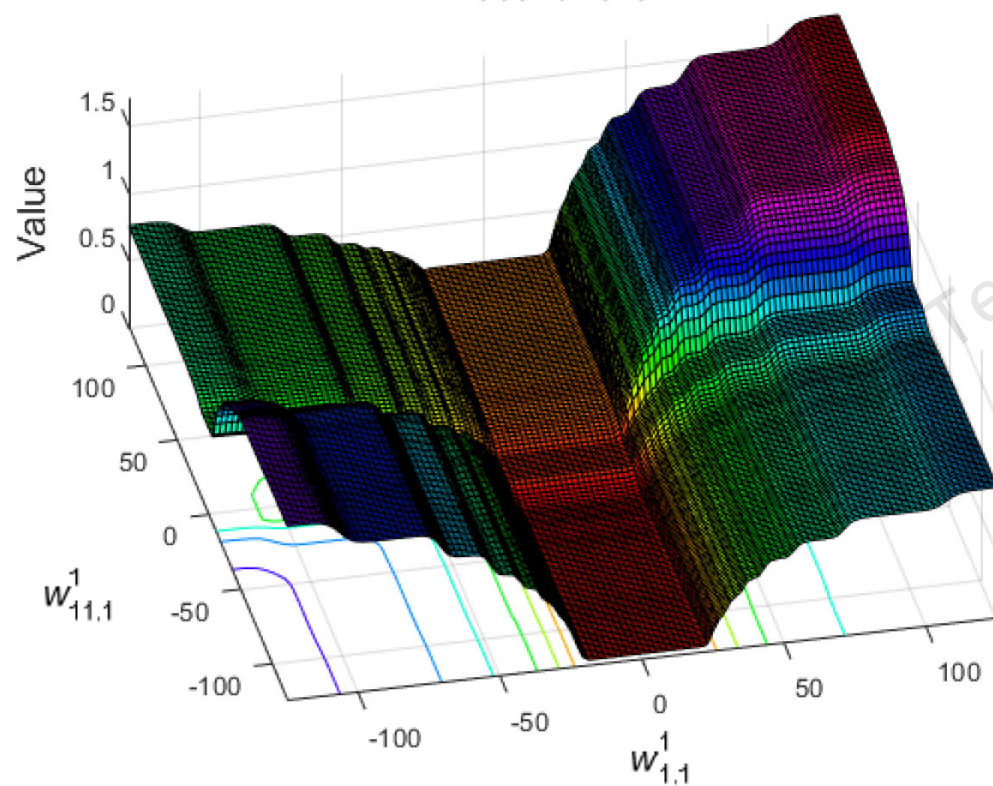
The middle terrain also has many special points with an integer derivative close to zero.

FBP can converge to the global optimal point, and its convergence speed is **very fast**.

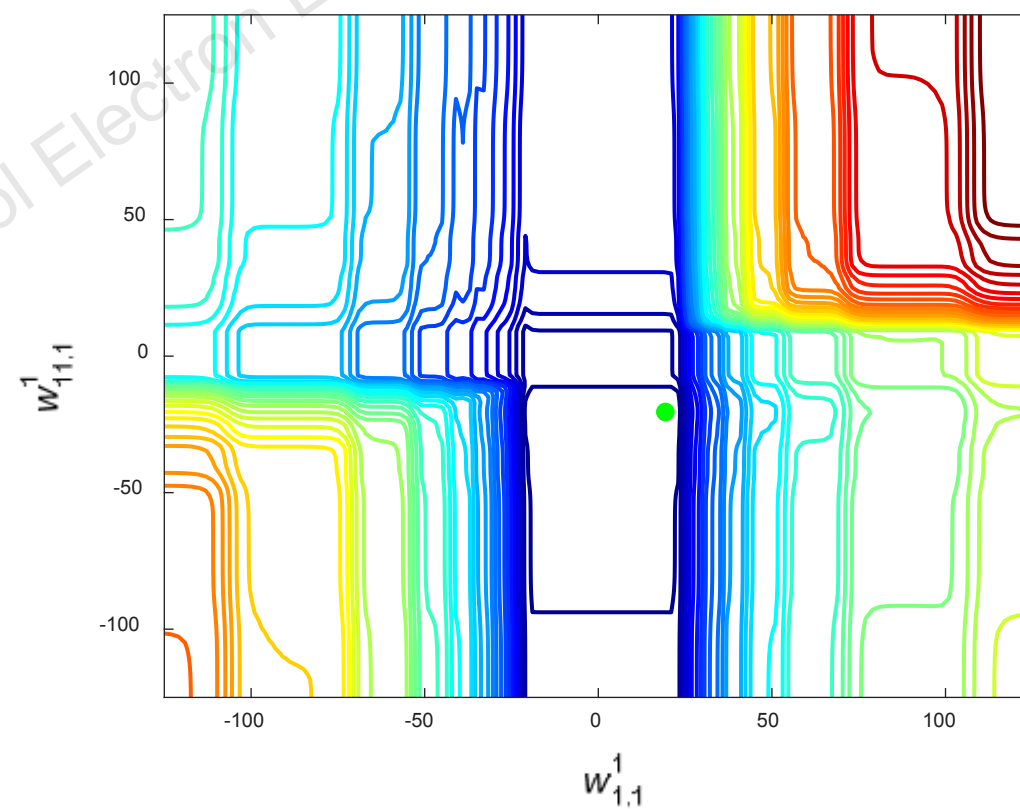
Stable running results

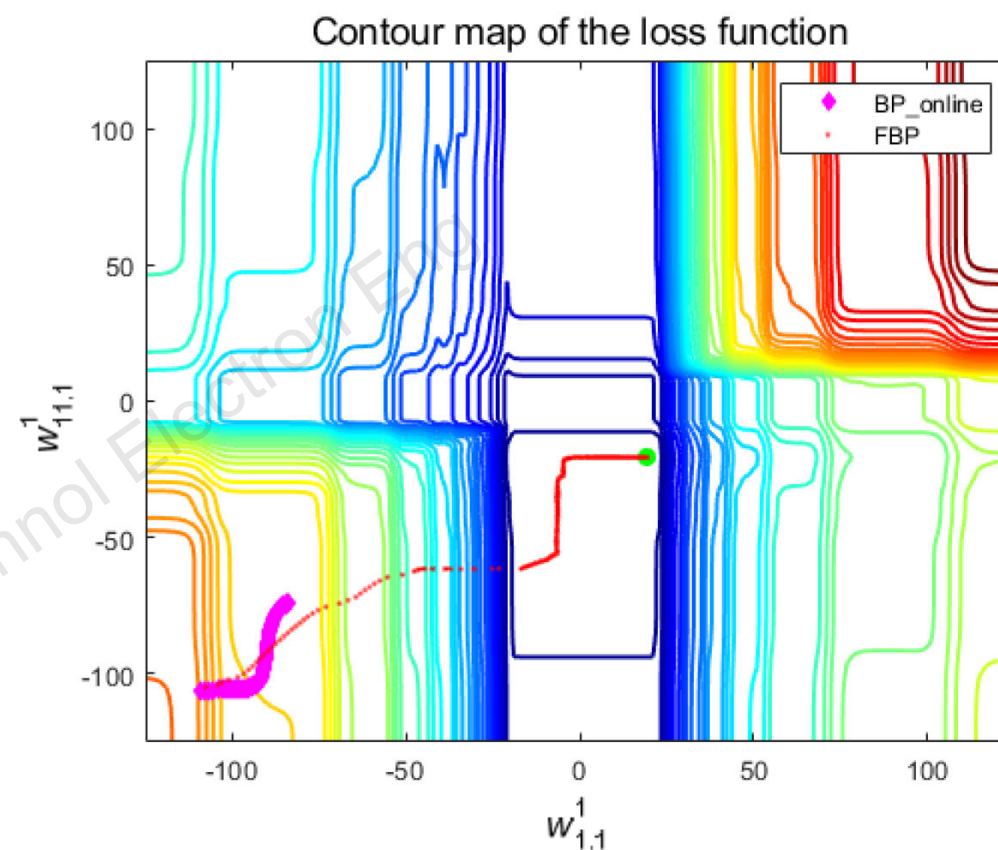
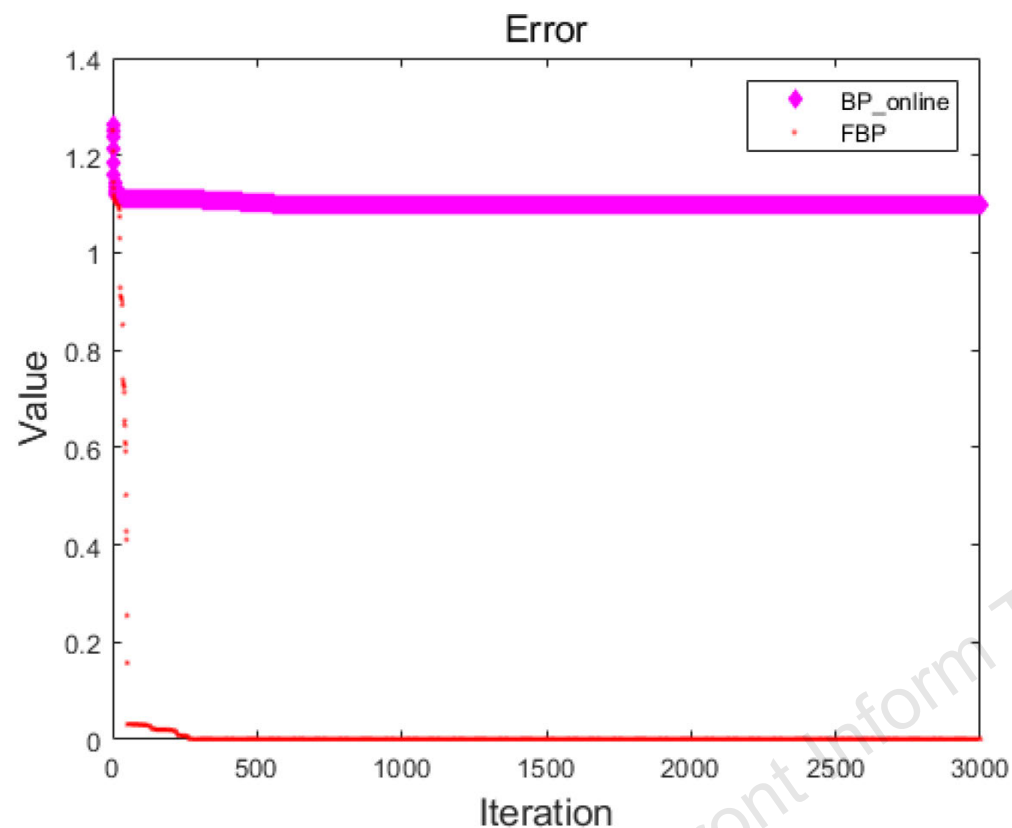
Example 2

Loss function



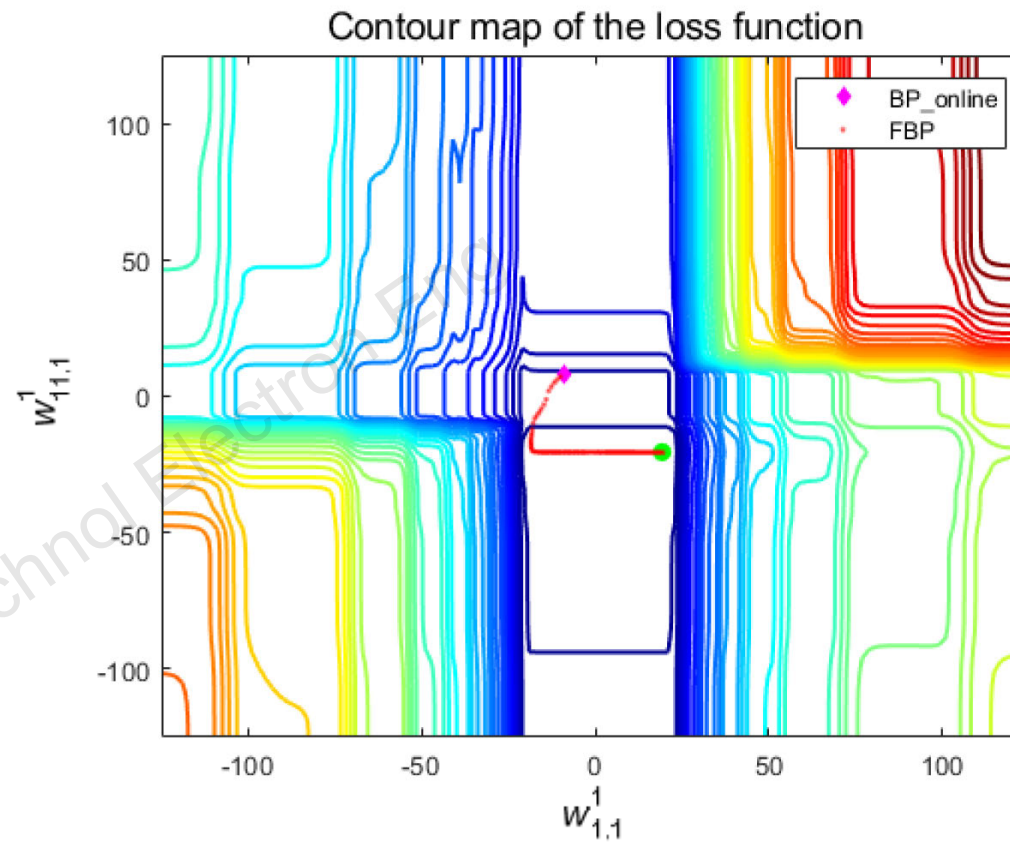
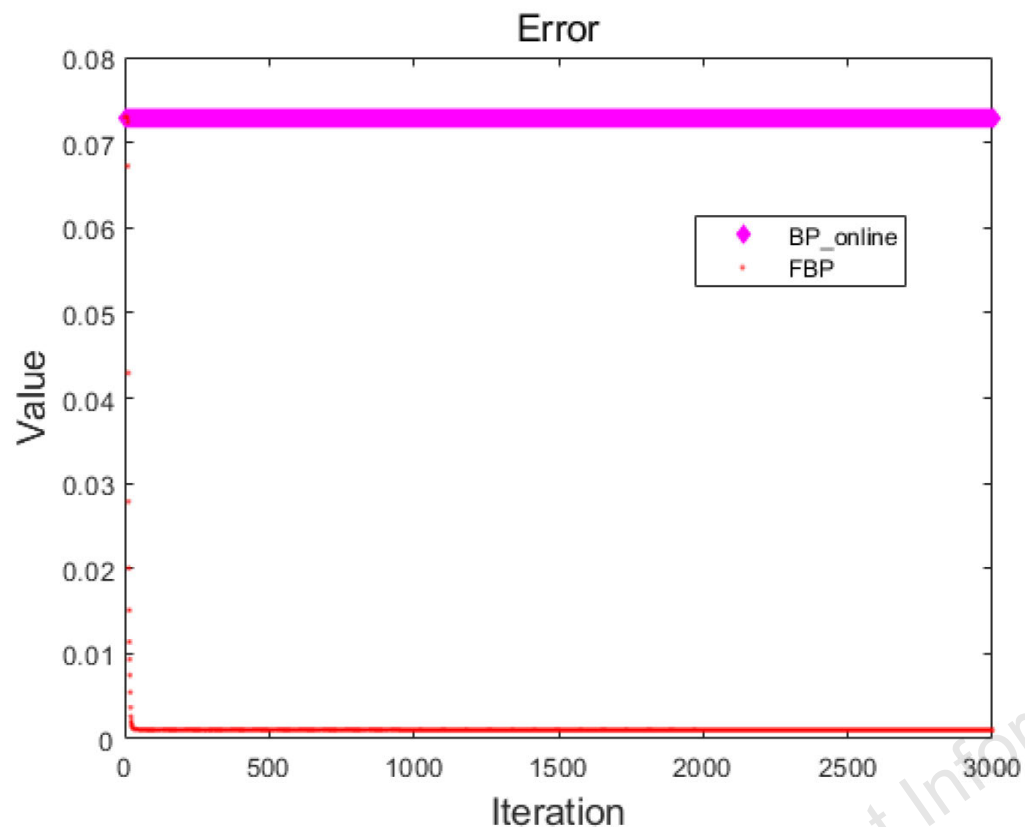
Contour map of the loss function





BPNN cannot converge to the global optimal point.

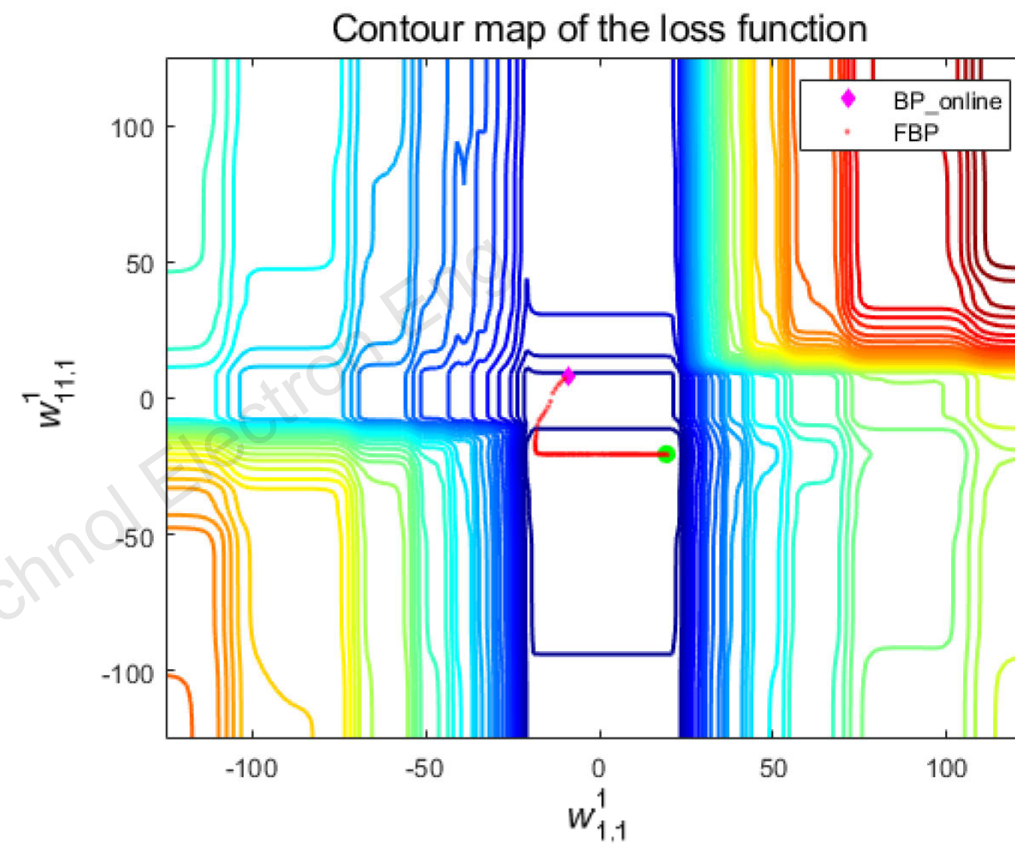
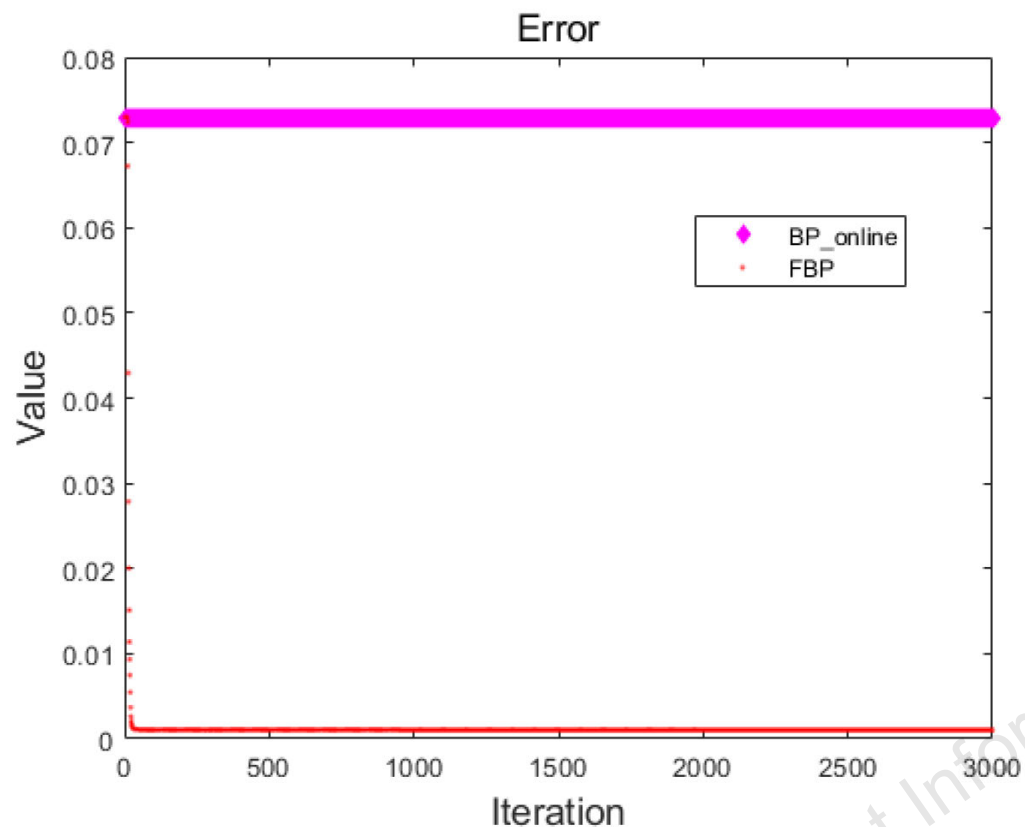
FBPNN can converge to the global optimal point, and its convergence speed is very fast.



Specifically, the initial point is just on the **local optimum point**.

Gradient does NOT disappear

Stable running results



Even if the global minimum is **close** to 0, FBPNN can converge to the **global optimal point**, and its convergence speed is **very fast**.



四川大学
SICHUAN UNIVERSITY

计算机学院



海纳百川，有容乃大

以人为本，崇尚学术

PART 4

Conclusions

Front Inform Technol Electron Eng



Simple

+

Valid in convex functions

+

Valid in non-convex functions

+

Advantages

Small computational burden

+

Small storage space

+

Fewer parameters

+

Stable running results

+

Gradient does not disappear

+

Stronger ability to express information

+

Fast

+

Even if the minimum value of the function is close to 0, FBPNN can also be valid.





四川大学
SICHUAN UNIVERSITY

计算机学院

Yi-fei PU (蒲亦非) Email: puyifei@scu.edu.cn

Yi-fei PU is a full professor and doctoral supervisor with the College of Computer Science, Sichuan University. He was supported by the Thousand Talents Program of Sichuan Province and elected as one of the Academic and Technical Leaders of Sichuan Province. He focuses on the application of fractional calculus to signal processing, image processing, circuits and systems, and machine intelligence. He has published as the first author about 25 papers in SCI-indexed journals such as *INT J NEURAL SYST*, *IEEE TIP*, *IEEE TNNLS*, *IEEE TCS I*, *IEEE IoT-J*, *IEEE ITSM*, *IEEE IS*, *IEEE/CAA JAS*, *Sci China Ser F*, and *Sci China Inform Sci*. He held several research projects, such as the National Key Research and Development Program of China, National Natural Science Foundation of China, and Returned Overseas Chinese Scholars Project of the Ministry of Education of China, and holds 14 China patents of invention as the first or single inventor on fractional-order signal processing.



中國石油大學 (华东)
CHINA UNIVERSITY OF PETROLEUM

Jian WANG (王健) Email: wangjiannl@upc.edu.cn

Jian WANG is currently an associate professor and serves as Head of the Laboratory for Intelligent Information Processing with the College of Science, China University of Petroleum (East China), Qingdao, China. Prof. WANG was a recipient of the Natural Science Academic Achievement Award of Liaoning Province in 2012 and 2013. His research interests include neural networks, machine learning, and data mining. He has served as General Chair, Program Chair, and Co-program Chair of several conferences such as the International Symposium on New Trends in Computational Intelligence, IEEE Symposium Series on Computational Intelligence, and International Symposium on Neural Networks. He is serving as an Editor of *Neural Computing and Applications* and an Associate Editor of *Journal of Applied Computer Science Methods* and *IEEE Transactions on Neural Networks and Learning Systems*.