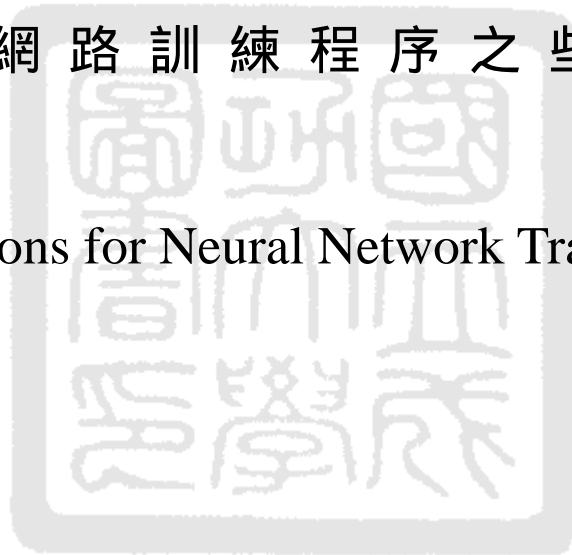


國立成功大學
航空太空工程研究所
碩士論文

類神經網路訓練程序之些許建議

A Few Suggestions for Neural Network Training Algorithm



研究生：陳威昇
指導教授：胡潛濱

中華民國九十二年六月

國立成功大學 碩士論文

類神經網路訓練程序之些許建議

A Few Suggestions for Neural Network

Training Algorithm

研究生：陳威昇

本論文業經審查及口試合格特此證明

論文考試委員

吳光鐘 吳光鐘

馬劍清 馬劍清

許梅娟 許梅娟

江達雲 江達雲

胡潛濱 胡潛濱

指導教授：胡潛濱 胡潛濱

所 長：趙怡欽 趙怡欽

中華民國 九十二年 六 月 二十 日

摘要

題 目：類神經網路訓練程序之些許建議

研 究 生：陳威昇

指導教授：胡潛濱

倒傳遞網路被廣泛的應用在各個領域，倒傳遞網路通常都採用最陡坡降法來當學習法則。但是，最陡坡降法在使用上有著許多的缺點，例如其網路所需的訓練時間太過沉長，以及容易掉入局部最小值等問題。

本文藉由五種訓練法則(最陡坡降法、共軛梯度法、LM Method、HWO-OWO Method、CGLM Method)去訓練三種類型的例子及一個實際應用例題，從中觀察每種方法的優缺點，並嘗試歸納出一些有用的建議。其中，CGLM Method 是將共軛梯度法及 LM Method 合併使用，希望能比其他方法擁有更佳的收斂性。

ABSTRACT

Subject : A Few Suggestions for Neural Network Training Algorithm

Student : Wei-Sheng Chen

Advisor : Chyanbin Hwu

Back-propagation neural network is widely applied to every field. It usually adopts the steepest descent method to search minimum of objective function. But the steepest descent method requires lengthy training time and is easy to trap into local minimum.

In order to speed up the convergence, this text uses four different training methods(conjugate gradient method, Levenberg Marquardt method, HWO-OWO method and conjugate gradient-Levenberg Marquardt method) for back-propagation neural network. We observe the shortcomings and advantages of every method and then try to induce some suggestions. Conjugate gradient-Levenberg Marquardt method is to combine the conjugate gradient method and Levenberg Marquardt method. The purpose of this method is to possess better convergence than the other methods.

目 錄

摘要

英文摘要

誌謝

目錄.....	i
表目錄.....	iii
圖目錄.....	iv
符號說明.....	vii

第一章 緒論

1.1 前言.....	1
1.2 文獻回顧.....	1
1.3 研究動機.....	3

第二章 類神經網路

2.1 簡介.....	4
2.2 功能.....	5
2.3 分類.....	6
2.4 架構.....	8

第三章 倒傳遞網路

3.1 順向傳遞規則.....	10
3.2 逆向回傳學習法則---最陡坡降法.....	12
3.3 逆向回傳學習法則---共軛梯度法.....	13
3.4 逆向回傳學習法則---LM Method.....	15
3.5 逆向回傳學習法則---HWO-OWO method.....	24
3.5.1 OWO.....	25

3.5.2 HWO.....	26
3.6 逆向回傳學習法則---CGLM Method.....	28
第四章 範例結果與實際應用	
4.1 基本測試題訓練結果.....	31
4.2 基本函數訓練結果.....	31
4.3 混和型訓練結果.....	33
4.4 實際例題訓練結果.....	34
4.5 結果與討論.....	37
第五章 結論與建議	
5.1 結論.....	39
5.2 建議.....	39
參考文獻.....	41
附錄.....	43
附表.....	46
附圖.....	63

自述

著作權聲明

表 目 錄

表 4.1	類型一 基本測試題.....	46
表 4.2	類型二 基本函數題.....	47
表 4.3	類型三(a) 二對一題型.....	49
表 4.4	類型三(b) 一對二題型.....	51
表 4.5	類型三(c) 二對二題型.....	53
表 4.6	類型三(d) 二對一混合題型.....	55
表 4.7	訓練對輸入資料.....	56
表 4.8	訓練對輸出資料(機翼展弦比等於二).....	58
表 4.9	訓練對輸出資料 (機翼展弦比等於四)	60
表 4.10	機翼展弦比等於二之測試結果.....	62
表 4.11	機翼展弦比等於四之測試結果.....	62

圖 目 錄

圖 2.1	生物神經元.....	63
圖 2.2	人工神經元.....	63
圖 2.3	類神經網路學習程序.....	63
圖 2.4	轉換函數種類.....	64
圖 2.5	類神經網路架構.....	65
圖 3.1	不含轉換函數之目標函數.....	66
圖 3.2	CGM Method 流程圖.....	67
圖 4.1	訓練結果 - 類型一 EX1 (2-10-1).....	68
圖 4.2	訓練結果 - 類型一 EX2 (2-5-1).....	69
圖 4.3	訓練結果 - 類型一 EX3 (2-2-1).....	70
圖 4.4	訓練結果 - 類型二 EX1 (1-8-1).....	71
圖 4.5	訓練結果 - 類型二 EX1 (1-20-1).....	72
圖 4.6	訓練結果 - 類型二 EX2 (1-3-1).....	73
圖 4.7	訓練結果 - 類型二 EX2 (1-20-1).....	74
圖 4.8	訓練結果 - 類型二 EX3 (1-20-1).....	75
圖 4.9	訓練結果 - 類型二 EX4 (1-20-1).....	75
圖 4.10	訓練結果 - 類型二 EX5 (1-20-1).....	76
圖 4.11	訓練結果 - 類型二 EX6 (1-20-1).....	76
圖 4.12	訓練結果 - 類型二 EX7 (1-20-1).....	77
圖 4.13	訓練結果 - 類型二 EX8 (1-20-1).....	77
圖 4.14	訓練結果 - 類型二 EX9 (1-20-1).....	78
圖 4.15	訓練結果 - 類型二 EX10 (1-20-1).....	78
圖 4.16	訓練結果 - 類型二 EX11 (1-20-1).....	79
圖 4.17	訓練結果 - 類型二 EX12 (1-20-1).....	79
圖 4.18	訓練結果 - 類型二 EX13 (1-20-1).....	80
圖 4.19	訓練結果 - 類型二 EX14 (1-20-1).....	80
圖 4.20	訓練結果 - 類型三(a) EX1 (2-20-1).....	81
圖 4.21	訓練結果 - 類型三(a) EX2 (2-20-1).....	81

圖 4.22	訓練結果 - 類型三(a) EX3 (2-20-1).....	82
圖 4.23	訓練結果 - 類型三(a) EX4 (2-20-1).....	82
圖 4.24	訓練結果 - 類型三(a) EX5 (2-20-1).....	83
圖 4.25	訓練結果 - 類型三(a) EX6 (2-20-1).....	83
圖 4.26	訓練結果 - 類型三(a) EX7 (2-20-1).....	84
圖 4.27	訓練結果 - 類型三(a) EX8 (2-20-1).....	84
圖 4.28	訓練結果 - 類型三(a) EX9 (2-20-1).....	85
圖 4.29	訓練結果 - 類型三(a) EX9 (2-50-1).....	85
圖 4.30	訓練結果 - 類型三(a) EX10 (2-20-1)	86
圖 4.31	訓練結果 - 類型三(a) EX10 (2-50-1)	86
圖 4.32	訓練結果 - 類型三(a) EX11 (2-20-1)	87
圖 4.33	訓練結果 - 類型三(a) EX12 (2-20-1)	87
圖 4.34	訓練結果 - 類型三(b) EX1 (1-8-2)	88
圖 4.35	訓練結果 - 類型三(b) EX2 (1-8-2)	88
圖 4.36	訓練結果 - 類型三(b) EX3 (1-3-2)	89
圖 4.37	訓練結果 - 類型三(b) EX4 (1-8-2)	89
圖 4.38	訓練結果 - 類型三(b) EX5 (1-8-2)	90
圖 4.39	訓練結果 - 類型三(b) EX6 (1-8-2)	90
圖 4.40	訓練結果 - 類型三(b) EX7 (1-8-2)	91
圖 4.41	訓練結果 - 類型三(b) EX8 (1-8-2)	91
圖 4.42	訓練結果 - 類型三(b) EX9 (1-8-2)	92
圖 4.43	訓練結果 - 類型三(b) EX9 (1-20-2)	92
圖 4.44	訓練結果 - 類型三(b) EX10 (1-40-2)	93
圖 4.45	訓練結果 - 類型三(b) EX10 (1-60-2)	93
圖 4.46	訓練結果 - 類型三(c) EX1 (2-50-2).....	94
圖 4.47	訓練結果 - 類型三(c) EX2 (2-40-2).....	94
圖 4.48	訓練結果 - 類型三(c) EX3 (2-20-2).....	95
圖 4.49	訓練結果 - 類型三(c) EX4 (2-40-2).....	96
圖 4.50	訓練結果 - 類型三(c) EX4 (2-50-2).....	96

圖 4.51	訓練結果 - 類型三(c) EX5 (2-20-2).....	97
圖 4.52	訓練結果 - 類型三(c) EX6 (2-50-2).....	97
圖 4.53	訓練結果 - 類型三(c) EX7 (2-20-2).....	98
圖 4.54	訓練結果 - 類型三(c) EX7 (2-50-2).....	98
圖 4.55	訓練結果 - 類型三(c) EX8 (2-40-2).....	99
圖 4.56	訓練結果 - 類型三(c) EX8 (2-50-2).....	99
圖 4.57	訓練結果 - 類型三(d) EX1 (2-20-1).....	100
圖 4.58	訓練結果 - 類型三(d) EX1 (2-50-1).....	100
圖 4.59	訓練結果 - 類型三(d) EX2 (2-20-1).....	101
圖 4.60	訓練結果 - 類型三(d) EX2 (2-40-1).....	101
圖 4.61	訓練結果 - 類型三(d) EX3 (2-20-1).....	102
圖 4.62	訓練結果 - 類型三(d) EX3 (2-50-1).....	102
圖 4.63	訓練結果 - 類型三(d) EX4 (2-8-1).....	103
圖 4.64	訓練結果 - 類型三(d) EX4 (2-20-1).....	103
圖 4.65	訓練結果 - 類型三(d) EX5 (2-8-1).....	104
圖 4.66	訓練結果 - 類型三(d) EX5 (2-20-1).....	104
圖 4.67	訓練結果 - 類型三(d) EX6 (2-20-1).....	105
圖 4.68	訓練結果 - 類型三(d) EX6 (2-50-1).....	105
圖 4.69	訓練結果 - 類型三(d) EX7 (2-20-1).....	106
圖 4.70	訓練結果 - 類型三(d) EX7 (2-50-1).....	106
圖 4.71	訓練結果 - 類型三(d) EX8 (2-8-1).....	107
圖 4.72	訓練結果 - 類型三(d) EX8 (2-40-1).....	107
圖 4.73	訓練結果 - 實例-機翼展弦比等於 2 (5-12-5).....	108
圖 4.74	訓練結果 - 實例-機翼展弦比等於 4 (5-12-5).....	109
圖 4.75	訓練結果 - 實例-機翼展弦比等於 2 (5-25-5).....	110
圖 4.76	訓練結果 - 實例-機翼展弦比等於 4 (5-25-5).....	110
圖 4.77	訓練結果 - 實例-機翼展弦比等於 2 (5-30-5).....	111
圖 4.78	訓練結果 - 實例-機翼展弦比等於 4 (5-30-5).....	111

符 號 說 明

\mathbf{b}_i	第 i 層之偏權值
c_1, c_2	調整 μ 大小
\mathbf{d}	目標值
\mathbf{D}	定義於(3.70)
\mathbf{e}	類似權重的性質，在 HWO 中更新權重用
E	目標函數
$\nabla \mathbf{E}$	目標函數對權重的一次微分
f	轉換函數
f', f''	轉換函數的一次及二次微分
\mathbf{F}'_i	第 i 層之轉換函數的一次微分
\mathbf{g}_i	第 i 層之權重與輸入乘積值
\mathbf{h}_i	第 i 層之輸入
i, j, k	用來表示某參數的所在位置
\mathbf{J}	Hessian Matrix
m	表最終輸出層
n	訓練對的個數
\mathbf{p}	訓練對

S	搜尋方向
\mathbf{W}_i	第 i 層之權重
w_{ijk}	第 i 層之權重中的某一元素
$\Delta \mathbf{W}$	權重修正量
$\Delta \mathbf{W}^-$	前次權重修正量
\mathbf{y}	網路輸出向量
α	慣性因子
β	定義於(3.90)及(3.91)
γ	定義於(3.85)及(3.86)
δ	局部梯度
η	學習速率
μ	用來調整 LM Method 的搜尋方向
μ^-	前次疊代之 μ

第一章

緒論

1.1 前言

不可否認的，現今電腦擁有比人腦更快速的計算能力，而且所得的結果具有高度的精準性及可靠性，這是人腦所無法比的，不過還是有許多的工作人腦卻比電腦優秀的多，例如：手寫文字辨識、使用者無關語音辨識、專業決策工作等。(黃威龍，2002)

人腦還有一個比辨識能力更重要的能力 - 學習能力。這是人類能賴以生存與適應環境的重要因素。一般的電腦並不具備此能力，就算近幾年所發展出一些人工智慧產品，也遠比人腦遜色，為了將人腦與電腦的優點合而為一，於是類神經網路便由此而生了。

類神經網路是模擬簡化的生物神經網路，也因此具有學習的功能，擁有如此強大的功能，使類神經網路被廣泛的應用到各個領域，例如：工程故障診斷、最佳化問題求解、股價預測、財務分析、醫學疾病診斷影像辨識、語音辨識等。(葉宜成，1993)

1.2 文獻回顧

類神經網路發展至今約有 50 年，其最早出現的第一種類神經網路模式為 - 感知器(Perceptron)，是由 Rosenblatt(1957)所提出，但因

感知器缺乏隱藏層的學習演算法，所以其學習能力受到很大的限制，也由於沒有隱藏層的感知器無法解決 exclusive OR(簡稱 XOR)問題。這問題直到 1985 年，倒傳遞網路(Back-Propagation Neural Network)被提出才得以解決。而倒傳遞網路是在 1974 年由 Werbos 的博士論文中提出了隱藏層的學習法則，但這重大的進展在當時並沒有引起重視，直到 1985 年 Parker 再次提出倒傳遞網路，同年由 Rumelhart 等人共同發表一篇倒傳遞網路文章才使其廣為人知。(葉宜成，1993)

倒傳遞網路一般採用最陡坡降法(Steepest Descent Method)作為訓練網路的方法，在效率上有收斂緩慢且易收斂到局部最小值(Local Minima)的缺點，也因此許多不同的學習法則陸續被提出，先是被提出可調整的學習速率及增加慣性項(Yam and Chow, 1993)，另外，也有許多學者使用共軛梯度法來增加收斂速度(Charalambous, 1992; Towsey, et. al. 1995)，然而，為了更快速的收斂，便有學者把最陡坡降法和牛頓法優點合併之後的Marquardt法則拿來當倒傳遞網路的學習法則(Hagan and Menhaj, 1994; Fun and Hagan, 1996)，近年由Chen et. al. (1999)提出，將學習法則分成兩個部分處理，一是對輸出層做最佳化(OWO)，一是對其他的隱藏層做最佳化(HWO)，以解線性方程式的方式取代以最佳化疊代的過程，大大減少疊代的次數(Chen, et. al. 1999; Scalero and Tepedelenlioglu, 1992)。而也有學者針對倒傳遞

網路容易收斂到局部最小值的問題加以修正(Cho and Chow, 1999; Liang and Feng, 2002)，更有學者直接採用基因演算法來當倒傳遞網路的學習法則(Chen and Liao, 1998)。

1.3 研究動機

類神經網路中，應用最廣泛的應屬倒傳遞網路，但是，類神經網路最讓人詬病的是其需要一段不算短的時間來訓練網路，所以，從提出倒傳遞網路至今，不斷的有學者提出改善的方法，其方法大致上可以分為兩類，一是從根本的最佳化方法著手，另一是增加或修改最佳化的參數。

而在使用倒傳遞網路時，存在著許多不確定因素，例如要用幾個神經元？要用哪種的轉換函數？學習速率要設多少等等。不過，最重要的應該是，哪種題目要用哪種學習法則。而本文則藉由以不同的訓練法則，去訓練大量的例子，進而從中歸納出一些有用的建議。

第二章

類神經網路

2.1 簡介

要知道類神經網路如何產生，就要先了解生物神經網路如何運作，因為，類神經網路是人類製造出來，用來模仿生物神經網路的資訊處理系統，也就是所謂“學習”的能力。

生物神經網路是由約 1000 億個神經細胞所組成，神經細胞又稱神經元(Neuron)，神經元是腦組織的基本單元(圖 2.1)。當外在環境由不同形式的資訊或能量訊號(例如，光，電，熱，聲音等的刺激)，藉由身體上的感官器官轉換成電的訊號後，便會進入神經細胞中，而在神經元內，此輸入訊號經由突觸(Synapse)間的內部電位變化後，透過樹突(Dendrites)傳送至細胞本體(Soma)，再由軸突(Axon)傳送至樹突，成為下一個神經元的輸入訊號。其中，突觸內的影響強度就是生物神經網路儲存資訊的地方，生物神經網路的學習就是在調節突觸的強度。所以我們就以人工神經元來模擬生物神經元(圖 2.2)，再由許多人工神經元，組成類神經網路。人工神經元又稱處理單元(Processing Element, PE)或節點(Node)。(羅華強，2001)

人工神經元的目的就是希望能夠跟生物神經一般運作，所以在結構上具有與一般神經元相似的模型，主要有幾個重要的單元：(施柏

屹 , 2000)

- 輸入信號：此處類似神經元的樹突（Dendrites），負責接受來自其他神經元的信號。
- 權值 (Weight) 與偏權值 (Bias)：此處則相似於神經元之突觸（Synapse），在此輸入信號經由權值的改變造成各個信號強化或抑制的效果。
- 轉換函數 (Transfer Function)：此函數之目的在於將類神經元所接收到的信號，透過類似神經細胞本體（Soma）所調變處理的功能，將其轉換輸出。

2.2 功能

當要研究或解決一個現實問題，若想掌握住所需解決的問題，必須要能夠對問題的對象作出明確的定義，此對象可稱為系統 (System) 或程序 (Process)，可利用數學技巧將此對象以一系列的數學方程式來完全描述並定義清楚，則可稱為此數學方程式為一個程序的數學模式。

數學模式大致上可分為兩種：基本模式 (Fundamental Model) 與經驗模式 (Empirical Model)。所謂的基本模式就是根據程序的基本物理、化學定律為基礎而獲得的一種模式，這種建模的方法常常因為程

序內部的機構過於複雜，或是其中存在著一些未知或不確定的參數而導致失敗或是實用上價值甚低。

所謂經驗模式就是在對所研究程序機構一無所知的情況下，將程序視作黑盒子(Black Box)，而僅僅憑藉程序的輸入與輸出數據，透過數學技巧來決定程序的模式。此數學技巧指的是回歸分析(Regression Analysis)，即對所取得的數據作曲線擬和(Curve Fitting)。

類神經網路就是一種經驗建模的工具，而它的確是以曲線擬和方式工作。然而，類神經網路與更古典的經驗統計建模方法之間存在著顯著的不同。類神經網路優於經驗統計模式之處如下(羅華強，2001)

- 類神經網路具有更強的過濾能力(Filtering Capacity)。
- 類神經網路具有適應性學習能力。
- 類神經網路是真正的多輸入多輸出系統(MIMO)。

2.3 分類

類神經網路的種類繁多，一般可按網路架構的不同，或按照學習方法的不同，或按照不同的應用領域而有不同的名稱。若依網路架構來區分，則可分為前饋式網路(Feed-Forward Network)，回饋式網路(Feed-Back Network)。在前饋式網路中，各層神經元僅與下一層次的

神經元相連結，且神經元的輸出訊號只向下一層次的神經元傳遞。而在回饋式網路中，該層次的神經元可以與前面層次的神經元相連結，而將輸出訊號往前回饋傳遞。

根據類神經網路的學習型態，基本上，可分為以下兩種網路類型：監督式(supervised)和非監督式(unsupervised)

1. 監督式學習網路(Supervised Learning Network)

所謂監督式學習法則，是將已知正確的輸入資料與輸出資料間的對應關係做為學習範例，因此在每一次的學習中，網路所得到的輸出資料都有一組正確的資料與之比對。在輸出層經過比對後得到學習誤差值，可將此誤差值從輸出層逐層回傳到內部的隱藏層中，網路則根據此回傳的誤差資訊來調整鏈結上的權值(圖2.3)，此學習過程可一再重複，直到誤差降到可接受的範圍，也就代表學習完成。理論上此學習完成的網路，即可充分反映出訓練範例中輸入資料與輸出資料的對應關係(沈宗澤，2001)。監督式的網路以大量範例的學習作為預測(predictions)、分類(classifications)及決策(decisions)之用。以下是監督式代表性及實用性的模式：(張吉甫，2002)

(1)倒傳遞網路(Back-Propagation Network)

(2)多層函數連結網路(Multilayer Functional-Link Network)

(3)廣義化回歸網路(Generalized Regression Neural Network)

(4)學習向量量化網路(Learning Vector Quantization Network)

(5)機率類神經網路(Probabilistic Neural Network)

其中，廣義化回歸網路和機率類神經網路都屬於徑向基網路(Radial Basis Network)。

2. 非監督式學習網路(Unsupervised Learning Network)

非監督式網路訓練之前，並沒有明確定義輸出如何，因此，其產生輸出的結果將無法預知，程式會將有共通點的資料作分類，自己決定它要何種結果，然後再觀察結果是否有用，或者使用者告知網路最大可分為多少類型，網路通常會將資料依類型作分類。以下是非監督式代表性及實用性的模式：(葉宜成，1993)

(1)自我組織映射圖網路(Self-Organizing Map)

(2)自適應共振理論網路(Adaptive-Resonance Theory Network)

網路模式的應用，是依問題領域的不同而作選擇，在這些模式中，以監督式倒傳遞網路應用範圍最為廣泛，使用結果最為成功，已被應用在各個領域，所以本研究以監督式倒傳遞網路為主。

2.4 架構

類神經網路中最基本的運算單元為人工神經元(Neuron)，其運算架構如圖2.2所示，輸入訊號與權值相乘之後，傳入人工神經元中，這些

權值代表著輸入訊號對神經元的刺激強度。人工神經元將相乘後的值加總，再加上人工神經元的偏權值(Bias)，產生一刺激訊號，此訊號經轉換函數(Transfer Function)作用後(圖2.4)，得到經人工神經元處理後的輸出訊號，再傳至相連結的下一層人工神經元上做為其輸入訊號。將數個性質相同的人工神經元並列組成網路中的同一層次，再將數層次串接，每個層次的人工神經元之間經由連結鏈(Connection)來傳遞訊號，如此形成類神經網路，整體的類神經網路架構如圖2.5。

類神經網路架構中，輸入層(Input Layer)的作用是接受外界的輸入訊號，然後經由連結鏈將訊號傳到網路內部的隱藏層(Hidden Layer)。這些隱藏層是類神經網路運作的主要層次，其結構及數目決定了類神經網路的功能及能力。經由隱藏層的運算後，訊號傳至輸出層(Output Layer)上，得到類神經網路對外界輸入訊號所產生的反應訊號。在類神經網路的學習(Training)過程中，主要的作用是要調整神經元連結鏈上的權值(Weight)。經由這些權值的調整，類神經網路可將資料庫中輸入資料與輸出資料的對應關係記憶下來，並可以經由回想(Recalling)的過程，將此對應關係重現，應用於工程問題上。(沈宗澤，2001)

第三章

倒傳遞網路

所謂倒傳遞就是將誤差倒傳回前面的隱藏層，此種現象主要是因為傳統的倒傳遞網路是採最陡坡降法(Gradient Descent Method)當學習法則，整理過後，正好有一個跟誤差有關的參數(,local gradient)會一層接一層的往回傳，所以才會稱倒傳遞網路。

基本上倒傳遞網路學習訓練是由兩部分組成即：訊號順向傳遞(forward pass)及誤差逆向回傳(backward pass)，網路順向傳遞時的權重是固定的，另一方面，逆向回傳的權重先是不變，等取到參數值時，再依所取得的參數及所選擇的學習法則來調整權重。所以順向傳遞時，不論選擇何種學習法則，過程、步驟都不變，所以此章節先介紹順向傳遞的規則，逆向回傳部分，再依不同的學習法則分段做介紹。

3.1 順向傳遞規則

假設一類神經網路，除輸入層外，隱藏層與輸出層合計有 m 層。若 \mathbf{p} 表輸入， \mathbf{y} 表輸出，令 $\mathbf{p} = \mathbf{h}_0$ 、 $\mathbf{y} = \mathbf{h}_m$ 。每一層的神經元個數為 S_i ， $i = 0, 1, 2, \dots, m$ ，則介於第 $i-1$ 層與第 i 層間所對應的權重為

$$\mathbf{W}_i = \begin{bmatrix} w_{i11} & w_{i12} & \cdots & w_{i1S_{i-1}} \\ w_{i21} & w_{i22} & \cdots & w_{i2S_{i-1}} \\ \vdots & \vdots & \ddots & \vdots \\ w_{iS_i1} & w_{iS_i2} & \cdots & w_{iS_iS_{i-1}} \end{bmatrix}, \quad i = 1, 2, \dots, m \quad (3.1)$$

若將偏權值併入考慮的話，則含偏權值的權重就變為

$$\mathbf{W}_i = \begin{bmatrix} w_{i11} & w_{i12} & \cdots & w_{i1S_{i-1}} & w_{i1S_{i-1}+1} \\ w_{i21} & w_{i22} & \cdots & w_{i2S_{i-1}} & w_{i2S_{i-1}+1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ w_{iS_i1} & w_{iS_i2} & \cdots & w_{iS_iS_{i-1}} & w_{iS_iS_{i-1}+1} \end{bmatrix}, \quad i=1,2,\dots,m \quad (3.2)$$

其中

$$\begin{bmatrix} w_{i1S_{i-1}+1} \\ w_{i2S_{i-1}+1} \\ \vdots \\ w_{iS_iS_{i-1}+1} \end{bmatrix} = \begin{bmatrix} b_{i1} \\ b_{i2} \\ \vdots \\ b_{iS_i} \end{bmatrix} = \mathbf{b}_i, \quad i=1,2,\dots,m \quad (3.3)$$

\mathbf{b}_i 為偏權值。為了對應權重的改變，所以輸入也要跟著變為

$$\mathbf{h}_i = \begin{bmatrix} h_{i1} \\ h_{i2} \\ \vdots \\ h_{iS_i} \\ 1 \end{bmatrix}, \quad i=0,1,2,\dots,m-1 \quad (3.4)$$

如此，每一層的輸出即可表示為

$$\mathbf{h}_i = f(\mathbf{W}_i \mathbf{h}_{i-1}), \quad i=1,2,\dots,m \quad (3.5)$$

從輸入層到輸出層輸出可表示為

$$\mathbf{y} = \mathbf{h}_m = f(\mathbf{W}_m f(\mathbf{W}_{m-1} f(\cdots f(\mathbf{W}_2 f(\mathbf{W}_1 \mathbf{h}_0)))) \quad (3.6)$$

其中， f 為轉換函數，在本文中，轉換函數選定為

$$f(x) = \frac{1}{1+e^{-x}} \quad (3.7)$$

目標函數等於

$$E = \frac{1}{2} \sum_{j=1}^{S_m} (d_j - h_{mj})^2 \quad (3.8)$$

式中 d_j 表輸出之目標值 (target value) 若為批次學習 (Batch Learning) , 則變為

$$E = \frac{1}{2} \sum_{p=1}^n \sum_{j=1}^{S_m} [(d_j)_p - (h_{mj})_p]^2 \quad (3.9)$$

其中, p 為訓練對 (Training Pattern) , $p = 1, 2, \dots, n$, n 為訓練對的個數。
 j 為輸出層中的某一神經元 , $j = 1, 2, \dots, S_m$ 。

3.2 逆向回傳學習法則---最陡坡降法

用最陡坡降法 (Steepest Descent Method) 來當學習法則是最普遍、最簡單的方法, 倒傳遞網路剛發展出來時, 即是採用最陡坡降法來當學習法則。最陡坡降法是沿著負梯度的方向去搜尋最佳值, 而最佳值通常落於斜率等於零的地方, 所以當搜尋到最佳值附近時, 其斜率變的很小, 因而造成權重的修正量也跟著變小, 以致影響收斂速度。

以(3.8)為目標函數, 依序對每一個權重微分, 所得到的微分值即為權重的修正量。

$$\Delta w_{ijk} = -\frac{\partial E}{\partial w_{ijk}} = \delta_{ij} h_{(i-1)k} \quad (3.10)$$

整理成矩陣形式, 可得下列通式

$$\Delta \mathbf{W}_i = -\nabla \mathbf{E}_i = \sum_{p=1}^n (\delta_i)_p (\mathbf{h}_{i-1}^T)_p \quad (3.11)$$

$$(\delta_i)_p = \begin{cases} (\mathbf{F}'_m)_p [(\mathbf{d})_p - (\mathbf{h}_m)_p], & i = m \\ (\mathbf{F}'_i)_p \mathbf{W}_{i+1}^T (\delta_{i+1})_p, & i = m-1, m-2, \dots, 2, 1 \end{cases} \quad (3.12)$$

其中，上標T表轉置矩陣， \mathbf{F}'_i 為轉換函數微分值所構成之對角矩陣

$$\mathbf{F}'_i = \begin{bmatrix} f'(g_{i1}) & 0 & \cdots & 0 \\ 0 & f'(g_{i2}) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & f'(g_{iS_i}) \end{bmatrix} \quad (3.13)$$

δ 為局部梯度(Local Gradient)，其定義為

$$\delta_i = \frac{\partial E}{\partial \mathbf{g}_i} \quad (3.14)$$

而 \mathbf{g}_i 表權重與其所對應的輸入乘積相加後的值

$$\mathbf{g}_i = \mathbf{W}_i \mathbf{h}_{i-1}, \quad i = 1, 2, \dots, m \quad (3.15)$$

由(3.11)式，如此便可得到更新後的權重值

$$\mathbf{W}_i = \mathbf{W}_i^- + \eta \Delta \mathbf{W}_i, \quad i = 1, 2, \dots, m \quad (3.16)$$

\mathbf{W}_i^- 表更新前的權重值。 η 為學習速率(learning rate)，這裡的學習速率就是在最佳化裡的步長(Step Length)，在本文中，學習速率一律設為1，因為，若要求出最佳步長，將會大幅拖長每次疊代的時間，而本文主要是比較每種方法之收斂性，所以每種方法之步長皆固定為1，並不失公平性。若加入慣性項(momentum term)則(3.16)式變為

$$\mathbf{W}_i = \mathbf{W}_i^- + \eta \Delta \mathbf{W}_i + \alpha \Delta \mathbf{W}_i^-, \quad i = 1, 2, \dots, m \quad (3.17)$$

α 為慣性因子，控制慣性項之比例， $0 \leq \alpha < 1$ 。 $\Delta \mathbf{W}_i^-$ 為前次疊代時的權重修正量。加上慣性項可以改善收斂過程中的振盪現象，並可增加網路的收斂速度。

3.3 逆向回傳學習法則---共軛梯度法

首先，定義出何謂共軛方向，設 \mathbf{A} 為一個 $n \times n$ 的對稱矩陣 (Symmetric Matrix)，若有一 n 個向量的向量集合 $\{S_i\}$ ，如果 $\{S_i\}$ 滿足下式

$$\mathbf{S}_i^T \mathbf{A} \mathbf{S}_j = 0, \quad \text{for all } i \neq j, i = 1, 2, \dots, n, j = 1, 2, \dots, n$$

則稱 $\{S_i\}$ 為一組 \mathbf{A} -共軛 (A-Conjugate) 的向量集合。

共軛梯度法 (Conjugate Gradient Method) 是沿著共軛方向來搜尋最佳值。使用共軛梯度法可以改善最陡坡降法在最佳值附近收斂緩慢的缺點。共軛梯度法與最陡坡降法的不同處只在於搜尋方向的差異而已，應用在類神經上可寫成 (Rao, 1996)

$$\Delta \mathbf{W}_i = -\nabla \mathbf{E}_i + \beta \Delta \mathbf{W}_i^-, \quad i = 1, 2, \dots, m \quad (3.18)$$

其中 $\nabla \mathbf{E}_i$ 為目標函數 E 依序對每個權重之偏微分，即如 (3.11) 所示。

β 的功用在於使新的權重修正量能與舊的權重修正量產生共軛，本文採用 Fletcher-Reeves Method (Rao, 1996)，所以 β 為

$$\beta_i = \frac{\nabla \mathbf{E}_i^T \nabla \mathbf{E}_i}{(\nabla \mathbf{E}_i^-)^T \nabla \mathbf{E}_i^-}, \quad i = 1, 2, \dots, m \quad (3.19)$$

其中， $\nabla \mathbf{E}_i$ 表目標函數的梯度， $\nabla \mathbf{E}_i^-$ 表示前次疊代目標函數的梯度。

第一次疊代時，

$$\Delta \mathbf{W}_i = -\nabla \mathbf{E}_i \quad (3.20)$$

當第二次以上時才用 (3.18) 式。更新後的權重值為

$$\mathbf{W}_i = \mathbf{W}_i^- + \eta \Delta \mathbf{W}_i \quad (3.21)$$

3.4 逆向回傳學習法則---Levenberg Marquardt Method

最陡坡降法在離最佳點尚遠時，擁有不錯的收斂性，但是在靠近最佳點時，因為梯度變小，所以收斂也跟著變緩慢。而牛頓法 (Newton's Method) 卻在靠近最佳點時擁有極佳的收斂性，因此 Levenberg Marquardt Method (以下簡稱 LM method) 主要的精神就是結合牛頓法跟最陡坡降法的優點，疊代初期所使用的搜尋方向較接近梯度方向，也就是最陡坡降法。疊代後期則趨近於牛頓法提供之方向。其控制方法是依靠參數 μ ，如下式所列

$$\tilde{\mathbf{J}} = \mathbf{J} + \mu \mathbf{I} \quad (3.22)$$

$$\mathbf{S} = -\tilde{\mathbf{J}}^{-1} \nabla E \quad (3.23)$$

其中， \mathbf{J} 為 Hessian Matrix， \mathbf{I} 為單位矩陣，而 \mathbf{S} 為搜尋方向， ∇E 為目標函數的一次微分。由 (3.22) 及 (3.23) 知，當 μ 足夠大時，LM method 會擁有最陡坡降法的特性；當 μ 較小時，LM method 會變成擁有牛頓法的特性，如此便可將兩種方法的優點結合在一起。

由於此時的 \mathbf{J} 及 ∇E 都不再分層考慮，而是將每一層每一個權重合併，一起考慮。所以本文將權重的排列順序規則定為從第 m 層到第 1 層，從最上面的神經元到最下面的神經元，如下

$$\begin{aligned}
w_{m11} &\Rightarrow w_1 \\
w_{m12} &\Rightarrow w_2 \\
&\vdots \\
w_{m1S_{m-1}} &\Rightarrow w_{S_{m-1}} \\
w_{m1(S_{m-1}+1)} &\Rightarrow w_{S_{m-1}+1}
\end{aligned}$$

此為與第 m 層第 1 個神經元相連接之權重排列順序，而最下面+1 是連偏權值一併考慮。下面繼續接第 m 層第 2 個神經元相連接之權重

$$\begin{aligned}
w_{m21} &\Rightarrow w_{S_{m-1}+2} \\
w_{m22} &\Rightarrow w_{S_{m-1}+3} \\
&\vdots \\
w_{m2(S_{m-1}+1)} &\Rightarrow w_{2 \times S_{m-1}+2}
\end{aligned}$$

箭頭右邊權重的下標表示編號第幾號；箭頭左邊權重的下標表示權重的所在位置。如此將所有的權重依序排列，則權重可表示成

$$\mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_{total} \end{bmatrix} \quad (3.24)$$

$$total = [S_m(S_{m-1} + 1)] + [S_{m-1}(S_{m-2} + 1)] + \cdots + [S_2(S_1 + 1)] \quad (3.25)$$

所以目標函數對權重一次微分可表示為

$$\nabla \mathbf{E} = \begin{bmatrix} \frac{\partial E}{\partial w_1} \\ \frac{\partial E}{\partial w_2} \\ \vdots \\ \frac{\partial E}{\partial w_{total}} \end{bmatrix} \quad (3.26)$$

而 Hessian Matrix 可表示為

$$\mathbf{J} = \begin{bmatrix} \frac{\partial^2 E}{\partial w_1 \partial w_1} & \frac{\partial^2 E}{\partial w_2 \partial w_1} & \cdots & \frac{\partial^2 E}{\partial w_{total} \partial w_1} \\ \frac{\partial^2 E}{\partial w_1 \partial w_2} & \frac{\partial^2 E}{\partial w_2 \partial w_2} & \cdots & \frac{\partial^2 E}{\partial w_{total} \partial w_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 E}{\partial w_1 \partial w_{total}} & \frac{\partial^2 E}{\partial w_2 \partial w_{total}} & \cdots & \frac{\partial^2 E}{\partial w_{total} \partial w_{total}} \end{bmatrix} \quad (3.27)$$

但是，要求得 \mathbf{J} 矩陣必須回歸原本的表示方式

$$\mathbf{J} = \frac{\partial^2 E}{\partial w_{ijk} \partial w_{i'j'k'}} \quad (3.28)$$

由(3.10)式可得出

$$\frac{\partial^2 E}{\partial w_{ijk} \partial w_{i'j'k'}} = \left(\frac{\partial \delta_{i'j'}}{\partial w_{ijk}} h_{(i'-1)k'} \right) \quad (3.29)$$

$i, i' = 1, 2, \dots, m$ ，表示位在第 i 層及位在第 i' 層； $j, j' = 1, 2, \dots, S_i$ ，表示位在第 i 層的第 j 個神經元及位在第 i' 層的第 j' 個神經元； $k, k' = 1, 2, \dots, S_{i-1} + 1$ ，表示位在第 $i-1$ 層的第 k 個神經元及位在第 $i'-1$ 層的第 k' 個神經元。 ijk 為相關且一同改變之參數； $i'j'k'$ 為相關且一同改變之參數， ijk 及 $i'j'k'$ 皆為描述權重所在位置，但是兩組參數各自獨立變化。

為了得知二階微分之變化情形，先取簡單的三層網路，每層只有一個神經元，且忽略偏權值，這樣方便觀察其規律性，而所對應的 Hessian Matrix 為

$$\mathbf{J} = \begin{bmatrix} \frac{\partial^2 E}{\partial w_3^2} & \frac{\partial^2 E}{\partial w_2 \partial w_3} & \frac{\partial^2 E}{\partial w_1 \partial w_3} \\ \frac{\partial^2 E}{\partial w_3 \partial w_2} & \frac{\partial^2 E}{\partial w_2^2} & \frac{\partial^2 E}{\partial w_1 \partial w_2} \\ \frac{\partial^2 E}{\partial w_3 \partial w_1} & \frac{\partial^2 E}{\partial w_2 \partial w_1} & \frac{\partial^2 E}{\partial w_1^2} \end{bmatrix} \quad (3.30)$$

代入(3.10)可將上式整理成

$$\mathbf{J} = - \begin{bmatrix} \frac{\partial(\delta_3 h_2)}{\partial w_3} & \frac{\partial(\delta_3 h_2)}{\partial w_2} & \frac{\partial(\delta_3 h_2)}{\partial w_1} \\ \frac{\partial(\delta_2 h_1)}{\partial w_3} & \frac{\partial(\delta_2 h_1)}{\partial w_2} & \frac{\partial(\delta_2 h_1)}{\partial w_1} \\ \frac{\partial(\delta_1 h_0)}{\partial w_3} & \frac{\partial(\delta_1 h_0)}{\partial w_2} & \frac{\partial(\delta_1 h_0)}{\partial w_1} \end{bmatrix} \quad (3.31)$$

因為是對稱矩陣，所以只需求出下三角矩陣即可，而下三角矩陣中之

h_1 ， h_0 對 w_3 或 w_2 之偏微分為零，因此

$$\frac{\partial(\delta_3 h_2)}{\partial w_3} = h_2 \frac{\partial \delta_3}{\partial w_3} ; \frac{\partial(\delta_2 h_1)}{\partial w_2} = h_1 \frac{\partial \delta_2}{\partial w_2} ; \frac{\partial(\delta_1 h_0)}{\partial w_1} = h_0 \frac{\partial \delta_1}{\partial w_1} \quad (3.32)$$

由此可知，只要 h 的下標數字比 δ 的下標數字還小， h 就可當成常數

處理。首先，求出位在最終輸出層的 δ_3 對 w_3 的微分

$$\frac{\partial \delta_3}{\partial w_3} = \frac{\partial}{\partial w_3} [f'(g_3) \cdot (d - h_3)] \quad (3.33)$$

$$= \left[\frac{\partial f'(g_3)}{\partial w_3} (d - h_3) + f'(g_3) \frac{\partial (d - h_3)}{\partial w_3} \right] \quad (3.34)$$

式中

$$\frac{\partial f'(g_3)}{\partial w_3} = \frac{\partial f'(g_3)}{\partial g_3} \frac{\partial g_3}{\partial w_3} = f''(g_3) h_2 \quad (3.35)$$

$$\frac{\partial(d-h_3)}{\partial w_3} = -\frac{\partial h_3}{\partial w_3} = -\frac{\partial f(g_3)}{\partial w_3} = -\frac{\partial f(g_3)}{\partial g_3} \frac{\partial g_3}{\partial w_3} = -f'(g_3)h_2 \quad (3.36)$$

所以

$$\frac{\partial \delta_3}{\partial w_3} = \gamma_3 h_2 \quad (3.37)$$

$$\gamma_3 = \left[-(f'(g_3))^2 + f''(g_3) \cdot (d - h_3) \right] \quad (3.38)$$

同理可得(推導過程詳見附錄)

$$\frac{\partial \delta_3}{\partial w_2} = \gamma_2 h_1 ; \quad \frac{\partial \delta_3}{\partial w_1} = \gamma_1 h_0 \quad (3.39)$$

其中

$$\gamma_2 = \gamma_3 w_3 f'(g_2) ; \quad \gamma_1 = \gamma_2 w_2 f'(g_1) \quad (3.40)$$

而 δ_2 對每一權重的偏微分為

$$\frac{\partial \delta_2}{\partial w_3} = f'(g_2) \left(\delta_3 + \frac{\partial \delta_3}{\partial w_3} w_3 \right) \quad (3.41)$$

$$\frac{\partial \delta_2}{\partial w_2} = w_3 \left(\frac{\partial f'(g_2)}{\partial w_2} \delta_3 + \frac{\partial \delta_3}{\partial w_2} f'(g_2) \right) \quad (3.42)$$

$$\frac{\partial \delta_2}{\partial w_1} = w_3 \left(\frac{\partial f'(g_2)}{\partial w_1} \delta_3 + \frac{\partial \delta_3}{\partial w_1} f'(g_2) \right) \quad (3.43)$$

其中

$$\frac{\partial f'(g_2)}{\partial w_2} = f''(g_2)h_1 = \beta_2^2 h_1 \quad (3.44)$$

$$\frac{\partial f'(g_2)}{\partial w_1} = f''(g_2)w_2 f'(g_1)h_0 = \beta_1^2 h_0 \quad (3.45)$$

而 δ_1 對每一權重的偏微分為

$$\frac{\partial \delta_1}{\partial w_3} = f'(g_1) w_2 \frac{\partial \delta_2}{\partial w_3} \quad (3.46)$$

$$\frac{\partial \delta_1}{\partial w_2} = f'(g_1) \left(\delta_2 + \frac{\partial \delta_2}{\partial w_2} w_2 \right) \quad (3.47)$$

$$\frac{\partial \delta_1}{\partial w_1} = w_2 \left(\frac{\partial f'(g_1)}{\partial w_1} \delta_2 + \frac{\partial \delta_2}{\partial w_1} f'(g_1) \right) \quad (3.48)$$

其中

$$\frac{\partial f'(g_1)}{\partial w_1} = f''(g_1) h_0 = \beta_1^1 h_0 \quad (3.49)$$

所以 Hessian Matrix 為

$$\mathbf{J} = \begin{bmatrix} \gamma_3(h_2)^2 & \text{symm.} \\ f'(g_2) \xi_3 h_1 & w_3 \varsigma_2 h_1 \\ w_2 f'(g_1) \frac{\partial \delta_2}{\partial w_3} h_0 & f'(g_1) \xi_2 h_0 & w_2 \varsigma_1 h_0 \end{bmatrix} \quad (3.50)$$

其中

$$\varsigma_r = \beta_r^r h_{r-1} \delta_{r+1} + \frac{\partial \delta_{r+1}}{\partial w_r} f'(g_r), \quad r = 1, 2 \quad (3.51)$$

$$\xi_s = \delta_s + \frac{\partial \delta_s}{\partial w_s} w_s, \quad s = 2, 3 \quad (3.52)$$

由上述可推論，如將網路擴大至 m 層，每一層依舊保持只有一個神經元，參考(3.37)至(3.40)，可推得

$$\frac{\partial \delta_m}{\partial w_i} = \gamma_i h_{i-1}, \quad i = 1, 2, \dots, m \quad (3.53)$$

$$\gamma_m = \left[-(f'(g_m))^2 + f''(g_m) \cdot (d - h_m) \right], \quad i = m \quad (3.54)$$

$$\gamma_i = \gamma_{i+1} w_{i+1} f'(g_i), \quad i = m-1, m-2, \dots, 2, 1 \quad (3.55)$$

觀察(3.37)、(3.41)、(3.46)及(3.39)、(3.42)、(3.47)可得知，需

利用 $\frac{\partial \delta_3}{\partial w_3}$ 、 $\frac{\partial \delta_3}{\partial w_2}$ 、 $\frac{\partial \delta_3}{\partial w_1}$ 才能求得 $\frac{\partial \delta_2}{\partial w_3}$ 、 $\frac{\partial \delta_2}{\partial w_2}$ 、 $\frac{\partial \delta_2}{\partial w_1}$ ，同理，需要有 δ_2 對

每一權重的偏微分，才能求得 δ_1 對每一權重的偏微分。推廣到 m 層

$$\frac{\partial \delta_{i'}}{\partial w_i} = \frac{\partial}{\partial w_i} [f'(g_{i'}) w_{i'+1} \delta_{i'+1}] \quad (3.56)$$

$$\frac{\partial w_{i'+1}}{\partial w_i} = \begin{cases} 1, i'+1 = i & i' = 1, 2, \dots, m-1 \\ 0, i'+1 \neq i & i = 1, 2, \dots, m \end{cases} \quad (3.57)$$

參考(3.44)、(3.45)、(3.49)可推得

$$\beta_{i'}^{i'} = f''(g_{i'}), \quad i = i' \quad (3.58)$$

$$\beta_i^{i'} = \beta_{i+1}^{i'} w_{i+1} f'(g_i), \quad i = i'-1, i'-2, \dots, 2, 1 \quad (3.59)$$

$$\frac{\partial f'(g_{i'})}{\partial w_i} = \beta_i^{i'} h_{i-1}, \quad i' = m-1, m-2, \dots, 2, 1 \quad (3.60)$$

所以 Hessian Matrix 為

J =

$$\begin{bmatrix} \gamma_m (h_{m-1})^2 & & & & \\ f'(g_{m-1}) \xi_m h_{m-2} & w_m \zeta_{m-1} h_{m-2} & & & \text{symm.} \\ w_{m-1} f'(g_{m-2}) \frac{\partial \delta_{m-1}}{\partial w_m} h_{m-3} & f'(g_{m-2}) \xi_{m-1} h_{m-3} & w_{m-1} \zeta_{m-2} h_{m-3} & & \\ \vdots & \vdots & \vdots & \ddots & \\ w_2 f'(g_1) \frac{\partial \delta_2}{\partial w_m} h_0 & w_2 f'(g_1) \frac{\partial \delta_2}{\partial w_{m-1}} h_0 & w_2 f'(g_1) \frac{\partial \delta_2}{\partial w_{m-2}} h_0 & \dots & w_2 \zeta_1 h_0 \end{bmatrix}$$

$$(3.61)$$

其中

$$\varsigma_r = \beta_r^r h_{r-1} \delta_{r+1} + \frac{\partial \delta_{r+1}}{\partial w_r} f'(g_r), \quad r = 1, 2, \dots, m-1 \quad (3.62)$$

$$\xi_s = \delta_s + \frac{\partial \delta_s}{\partial w_s} w_s, \quad s = 2, 3, \dots, m \quad (3.63)$$

再將網路擴大成每層有 S_i 個神經元，則 \mathbf{J} 可利用 (3.29) 式求出，觀察

(3.53)，(3.54)，(3.55) 可推得

$$\gamma_m^{mj'} = \begin{bmatrix} \gamma_{m1}^{mj'} \\ \gamma_{m2}^{mj'} \\ \vdots \\ \gamma_{mS_m}^{mj'} \end{bmatrix}, \quad i = m \quad (3.64)$$

其中，當 $i = m$ ， $j = j'$ 時

$$\gamma_{mj'}^{mj'} = \left[-(f'(g_{mj'}))^2 + f''(g_{mj'}) \cdot (d_{j'} - h_{mj'}) \right] \quad (3.65)$$

當 $i = m$ ， $j \neq j'$ 時， $\gamma_{mj'}^{mj'} = 0$ 。其中的 f' 和 f'' 為轉換函數的一次及二次

微分。而

$$\gamma_i^{mj'} = \mathbf{F}_i^T \mathbf{W}_{(i+1)}^T \gamma_{(i+1)}^{mj'}, \quad i = m-1, m-2, \dots, 2, 1 \quad (3.66)$$

位在輸出層的 δ_m 對每個權重的微分就可表示為，

$$\frac{\partial \delta_{mj'}}{\partial w_{ijk}} = \gamma_{ij}^{mj'} h_{(i-1)k} \quad (3.67)$$

$\delta_{mj'}$ 的下標 m 表示位在最終輸出層，也就是第 m 層；下標 j' 表示針對

位在最終輸出層 m 中的第 j' 個神經元。 $\gamma_{ij}^{mj'}$ 表示針對第 m 層第 j' 個神

經元而位在第 i 層第 j 個神經元之 γ 。有了輸出層之 δ 對每一權重的偏微分，就可利用(3.68)式依序求出 $m-1, m-2, \dots, 2, 1$ 層的 δ 對每一權重的偏微分，參考(3.56)，(3.57)，可推論出

$$\frac{\partial \delta_{i'}}{\partial w_{ijk}} = \frac{\partial}{\partial w_{ijk}} (\mathbf{F}_{i'}^T \mathbf{W}_{i'+1}^T \delta_{i'+1}), \quad i' = m-1, \dots, 2, 1 \quad (3.68)$$

而 $\frac{\partial \mathbf{W}_{i+1}^T}{\partial w_{ijk}}$ 則視分子和分母的權重所在位置是否相同，也就是下標是否

一致，如相同則其值為 1，若不同則為 0。如此便可一層接一層的求出 δ 對權重的偏微分值。參考(3.58)，(3.59)，(3.60)，可推論出

$$\beta_{i'}^{i'j'} = \begin{bmatrix} \beta_{i'1}^{i'j'} \\ \beta_{i'2}^{i'j'} \\ \vdots \\ \beta_{i'S_{i'}}^{i'j'} \end{bmatrix} \quad (3.69)$$

當 $i = i'$ 且 $j = j'$

$$\beta_{i'j'}^{i'j'} = f'(g_{i'j'}) \quad (3.70)$$

當 $i = i'$ 且 $j \neq j'$ ， $\beta_{i'j}^{i'j'} = 0$

$i' = m-1, m-2, \dots, 2, 1$ ， i' 表位在第 i' 層。 $j' = 1, 2, \dots, S_{i'}$ ， j' 表位在第 i' 層的第 j' 個神經元。 $\beta_{ij}^{i'j'}$ 表示針對第 i' 層第 j' 個神經元，且位在第 i 層第 j 個神經元之 β 。而 $\beta_i^{i'j'}$ 項則由下式求出

$$\beta_i^{i'j'} = \mathbf{F}_i^T \mathbf{W}_{(i+1)}^T \beta_{(i+1)}^{i'j'}, \quad i = i'-1, i'-2, \dots, 2, 1 \quad (3.71)$$

如此便可求出所需之轉換函數的一次微分 (f') 對每一個權重的偏微分

$$\frac{\partial f'(g_{i'j'})}{\partial w_{ijk}} = \beta_{ij}^{i'j'} h_{(i-1)k} \quad (3.72)$$

綜合上述式子，已可求出 \mathbf{J} 之下三角矩陣，由於 \mathbf{J} 為對稱，進而可求出完整的 Hessian Matrix，再代入(3.22)、(3.23)式，即可求出新權重

$$\mathbf{W} = \mathbf{W}^- - (\mathbf{J} + \mu \mathbf{I})^{-1} \nabla \mathbf{E} \quad (3.73)$$

LM Method 在大部分的情況下擁有比其他方法還好的收斂性，不過每疊代一次所花費的時間及記憶體也都比其他方法更多，在權重個數多時，將會造成一大缺點，且理論推導也較其他方法難，不過仍因其有高效率的收斂特性，故此方法仍舊常被拿來當倒傳遞網路的學習法則。

3.5 逆向回傳學習法則---HWO-OWO Method

此方法是由 Chen, et. al.(1999)所提出的新方法，其主要的構想是以解線性方程式來取代以往的疊代方式，並將權重的更新分為輸出層和隱藏層兩部分，而 OWO 是 Output Weight Optimization 的縮寫；HWO 是 Hidden Weight Optimization 的縮寫。

以往一般類神經網路的目標函數值都是以經過轉換函數作用後的輸出值再跟目標值作比較，而產生目標函數，然後再對目標函數選取一最佳化方法來進行權重的更新。而此方法最大的不同點為

1. 所採取的目標函數不再經過轉換函數的作用，網路在轉換函數前即可產生目標函數，也就是將網路輸出值及目標值對轉換函數作逆轉換 (Inverse)，如此便可將目標函數變為線性，避免掉去處理非線性的目標函數(圖 3.1)。

2. 網路中的每個神經元，皆有本身所對應的一個目標函數，也就是有幾個神經元就會有幾個目標函數，不再像以往，整個網路只有一個目標函數。

3. 權重的更新，是將每一個神經元所對應的目標函數，對此神經元所連接的權重取微分，並令其等於零，解聯立方程式，即可求得新的權重值。

3.5.1 OWO

以往的目標函數為

$$E = \frac{1}{2} \sum_{p=1}^n \sum_{j=1}^{S_m} [(d_j)_p - (h_{mj})_p]^2 \quad (3.74)$$

而在 OWO 的方法中，輸出層之每個神經元皆有本身的目標函數

$$E_{mj} = \frac{1}{2} \sum_{p=1}^n [(d_j^*)_p - (h_{mj}^*)_p]^2 \quad (3.75)$$

其中

$$d_j^* = f^{-1}(d_j) \quad (3.76)$$

$$h_{mj}^* = g_{mj} = \sum_{k=1}^{S_{m-1}+1} w_{mjk} h_{(m-1)k} = \mathbf{w}_{mj}^T \mathbf{h}_{m-1} \quad (3.77)$$

其中

$$\mathbf{w}_{mj}^T = \begin{bmatrix} w_{mj1} & w_{mj2} & \cdots & w_{mjS_{m-1}+1} \end{bmatrix} \quad (3.78)$$

目標函數對所連接的權重作微分

$$\frac{\partial E_{mj}}{\partial w_{mjk}} = - \sum_{p=1}^n \left[(d_j^*)_p - (h_{mj}^*)_p \right] \cdot (h_{(m-1)k})_p, \quad \begin{matrix} j = 1, 2, \dots, S_m \\ k = 1, 2, \dots, S_{m-1} + 1 \end{matrix} \quad (3.79)$$

令其等於零並將(3.77)之關係式代入可得

$$\begin{aligned} \sum_{p=1}^n (d_j^*)_p \cdot (h_{(m-1)k})_p &= \sum_{p=1}^n \left[\sum_{k=1}^{S_{m-1}+1} w_{mjk} (h_{(m-1)k})_p \right] (h_{(m-1)k})_p \\ &= \mathbf{w}_{mj}^T \sum_{p=1}^n (\mathbf{h}_{m-1})_p \cdot (h_{(m-1)k})_p \end{aligned} \quad (3.80)$$

先固定下標 j 再分別將 $k = 1, 2, \dots, S_{m-1} + 1$ 代入(3.80)式, 即可得到 $S_{m-1} + 1$

個方程式, 解 $S_{m-1} + 1$ 個聯立方程式, 得新的權重 \mathbf{w}_{mj}^T 。將 \mathbf{w}_{mj}^T 依

$j = 1, 2, \dots, S_m$ 排列, 即可組成 \mathbf{W}_m 。

$$\mathbf{W}_m = \begin{bmatrix} \mathbf{w}_{m1}^T \\ \mathbf{w}_{m2}^T \\ \vdots \\ \mathbf{w}_{mS_m}^T \end{bmatrix} \quad (3.81)$$

3.5.2 HWO

因為隱藏層中沒有所謂的目標值, 所以必須預估一個目標值

$$d_{ij}^* = \mathbf{w}_{ij}^T \mathbf{h}_{i-1} + \eta \delta_{ij} \quad (3.82)$$

其中

$$\mathbf{w}_{ij}^T = \begin{bmatrix} w_{ij1} & w_{ij2} & \cdots & w_{ijS_{i-1}+1} \end{bmatrix} \quad (3.83)$$

設隱藏層權重的更新為

$$\mathbf{w}_{ij}^+ = \mathbf{w}_{ij} + \eta \mathbf{e}_{ij} \quad (3.84)$$

其中

$$\mathbf{e}_{ij}^T = [e_{ij1} \quad e_{ij2} \quad \cdots \quad e_{ijS_{i-1}+1}] \quad (3.85)$$

故下次疊代網路的輸出為

$$h_{ij}^* = (\mathbf{w}_{ij}^T + \eta \mathbf{e}_{ij}^T) \mathbf{h}_{i-1} \quad (3.86)$$

所以隱藏層的目標函數變為

$$E_{ij} = \frac{1}{2} \sum_{p=1}^n [(d_{ij}^*)_p - (h_{ij}^*)_p]^2 \quad (3.87)$$

將(3.82)及(3.86)代入(3.87)，整理得

$$E_{ij} = \frac{1}{2} \eta^2 \sum_{p=1}^n [(\delta_{ij})_p - \mathbf{e}_{ij}^T (\mathbf{h}_{i-1})_p]^2 \quad (3.88)$$

目標函數對所連接的權重作微分

$$\frac{\partial E_{ij}}{\partial e_{ijk}} = -\eta^2 \sum_{p=1}^n [(\delta_{ij})_p - \mathbf{e}_{ij}^T (\mathbf{h}_{i-1})_p] \cdot (h_{(i-1)k})_p \quad (3.89)$$

令其等於零並整理

$$\begin{aligned} \sum_{p=1}^n (\delta_{ij})_p \cdot (h_{(i-1)k})_p &= \sum_{p=1}^n \left[\sum_{k=1}^{S_{i-1}+1} e_{ijk} (h_{(i-1)k})_p \right] (h_{(i-1)k})_p \\ &= \mathbf{e}_{ij}^T \sum_{p=1}^n (\mathbf{h}_{i-1})_p (h_{(i-1)k})_p \end{aligned} \quad (3.90)$$

$i=1,2,\dots,m-1$ ， i 表所在層數，先固定 i ，再固定 j ，再分別將 $k=1,2,\dots,S_{i-1}+1$ 代入(3.90)，即可得到 $S_{i-1}+1$ 個方程式，解 $S_{i-1}+1$ 個聯立方程式，得新的權重 \mathbf{e}_{ij}^T ，再代回(3.84)即可得到更新後的 \mathbf{w}_{ij}^T 。將 \mathbf{w}_{ij}^T

依 $j=1,2,\dots,S_i$ 排列，即可組成 \mathbf{W}_i 。

$$\mathbf{W}_i = \begin{bmatrix} \mathbf{w}_{i1}^T \\ \mathbf{w}_{i2}^T \\ \vdots \\ \mathbf{w}_{iS_m}^T \end{bmatrix} \quad (3.91)$$

3.6 逆向回傳學習法則---CGLM Method

在本文所謂的 CGLM Method，也就是把共軛梯度法及 LM Method 混合使用，簡稱 CGLM Method。因為共軛梯度法的共軛方向，不但共軛且含有梯度方向，所以疊代的初期依舊能快速收斂，更因為使用共軛方向而能快速接近局部最小值；而疊代後期則使用 LM Method，但此時參數的設定使 LM Method 趨近於牛頓法，這樣可以避免有時候 LM Method 的搜尋方向介於最陡坡降法和牛頓法之間，增加沒效率的疊代次數，所以本文提出此方法希望能縮短訓練時間。

其過程為(圖 3.2)，先使用共軛梯度法作為學習法則，直到誤差能收斂到 MSE(Mean Square Error)小於或等於 0.05 時，再轉為用 LM Method 當學習法則，直到收斂或者已達停止條件（疊代次數或者訓練時間）。若疊代中發生錯誤（如收斂到局部誤差）時，則重新給一個起始的權重，並回到共軛梯度法。

如(3.22)所述一般的 LM Method 使用參數 μ 來控制搜尋方向為偏向最陡坡降法或是牛頓法，再用 c_1 、 c_2 控制 μ 的大小，其條件為，

當誤差比前一次的誤差小的時後

$$\mu = c_1 \mu^- \quad (3.92)$$

當誤差比前一次的誤差大的時後

$$\mu = c_2 \mu^- \quad (3.93)$$

一般的 LM Method 參數 μ 起始值為 10000 , c_1 等於 0.25 , c_2 等於 2。

μ 起始值設為 10000 目的是要讓 LM Method 能在一開始訓練時偏向最陡坡降法 , μ^- 表上次疊代的 μ 值。合併法中 , 前半段的訓練已交給共軛梯度法 , 所以後半段主要是讓 LM Method 偏向牛頓法 , 故合併法中 μ 起始值為 0.00001 , 而 c_1 、 c_2 不變。如此 , 嚴格說來 , CGLM Method 應是共軛梯度法及牛頓法混合使用。

第四章

結果與討論

本章主要探討第三章所列各方法之優劣性。藉由舉出各種不同的函數加以訓練，並觀察每種方法所適用的函數或者是權重個數在某區間下適用何種方法，或者其他情況下，每種方法的適用性。所舉的範例分為三類：

1. 基本測試題---主要為類神經相關書籍中的示範例題，其所舉出的例題都為簡單且訓練對數目少(表 4.1)。
2. 基本函數 --- 此類為一些數學上常用的基本函數，如 $f(x) = \sin(x)$ ，其中 $f(x)$ 為目標值， x 為輸入值，此類例題全為一對一，也就是輸入向量和目標值向量的元素都只有一個(表 4.2)。
3. 混和型---此類型又細分為四種，主要是將基本函數作混和或相併，組合成一對多、多對一或多對多的形式(表 4.3，4.4，4.5，4.6)。

所訂定的標準為訓練時間，在所規定的訓練時間內無法收斂到容許誤差內的話，就算此方法在此題中無法收斂。這裡的時間是指電腦中央處理器運作的時間，而每一類型的例題會定不同的訓練時間，這是因為各種類型題目的訓練對個數並不相同，為求公平起見，不同的例題將會訂定不同的訓練時間。

4.1 基本測試題訓練結果

由圖 4.1, 4.2, 4.3 知, 在此類題中, 不管是疊代時間還是疊代次數都是以 OWO-HWO Method 這種方法最好。另外值得注意的是, 雖然 LM Method 及 CGLM Method 比最陡坡降法及共軛梯度法所需的疊代次數來得少, 但是, LM Method 及 CGLM Method 所花費的時間卻比其他兩種方法來的多。此外, 由圖 4.1, 4.3 可知, 共軛梯度法確實能改善最陡坡降法最後階段收斂緩慢的特性, 但是, 在 EX2 問題中(圖 4.2), 顯然共軛梯度法並不適用。

4.2 基本函數訓練結果

在此類問題中(圖 4.4 至 4.19。疊代次數標在每張圖右上方括弧內), 最陡坡降法及共軛梯度法都無法在所限定的時間內收斂。不過, 兩者的情形並不相同, 最陡坡降法都能出現收斂的趨勢, 但到最後總是因為收斂的曲線太平緩, 幾乎呈現一直線; 而共軛梯度法卻是一直來回的震盪, 完全沒有收斂的趨勢, 可能是因為沒有考慮到最佳步長的關係。

在所舉的 14 個例題中, HWO-OWO Method 有達到收斂標準的共有 4 題(4.5, 4.7, 4.8, 4.16)。本類題中, 隱藏層神經元數最多使用到 20 個, 觀察 HWO-OWO Method 有收斂的兩個例題可發現(圖

4.4, 4.5 及圖 4.6, 4.7), 隱藏層神經元較少時, HWO-OWO Method 並沒有收斂(圖 4.4 及圖 4.6), 要將神經元數加到足夠時, HWO-OWO Method 才會收斂(圖 4.5 及圖 4.7)。如果神經元數目不足的話, HWO-OWO Method 會和共軛梯度法一樣的震盪。為了證實是否將神經元往上加 HWO-OWO Method 就能收斂, 另外對 EX4, EX5, EX6, EX7, EX8 等五個沒收斂的例題加上神經元, 最多加到 100 個神經元, 其結果也是無法收斂。

而 LM Method 和 CGLM Method 在這類題中都擁有不錯的收斂性, 而這兩種方法依題目的不同, 有時 LM Method 比較快收斂(圖 4.4, 4.5, 4.15, 4.17, 4.18, 4.19), 有時卻是 CGLM Method 較快(圖 4.7, 4.8, 4.9, 4.10, 4.11, 4.12, 4.13, 4.14, 4.16), 由上面比較可知, 多數情形 CGLM Method 確實能比 LM Method 擁有更好的收斂性。

整體來說, 只看 HWO-OWO Method 有收斂的例題, HWO-OWO method 依舊能快速收斂, 就算需要比 LM Method 及 CGLM Method 更多的疊代次數, 但每次疊代所需的時間較少於 LM Method 及 CGLM Method, 最終所需的訓練時間還是比較短。不過, 就應用的廣泛性來說, LM Method 及 CGLM Method 能收斂的題數明顯多於 HWO-OWO Method。

4.3 混和型訓練結果

a 部分---在此部分中，最陡坡降法、共軛梯度法、HWO-OWO Method 皆無法收斂(圖 4.20 至 4.33)，不過最陡坡降法是收斂到局部最小值，而共軛梯度法、HWO-OWO Method 卻是反覆震盪。a 部分中，共有四個例題五種方法都無法收斂(圖 4.21, 4.28, 4.29, 4.30, 4.31, 4.32)，約三分之二的例題中 LM Method 及 CGLM Method 都能收斂，其中，CGLM Method 收斂優於 LM Method 的共有六題(圖 4.20, 4.22, 4.23, 4.26, 4.27, 4.33)。特殊題為圖 4.25，此題只有 LM Method 收斂而已。

b 部分---此部分只有一題沒收斂(圖 4.42 及 4.43)，其他九個例題至少都有一種方法有收斂。原因應該是此部分題型為一對二，所以訓練對組數較少，所以收斂較容易。CGLM Method 及 LM Method 各有四題領先，分別為圖 4.35, 4.37, 4.38, 4.45 以及圖 4.36, 4.39, 4.40, 4.41。而另一題較為特殊，只有 CGLM Method 有收斂而 LM Method 卻沒有(圖 4.34)。

c 部分---此部分因為是二對二類型，訓練對組數較多，目標函數較為複雜，所以共有五題全部的方法皆無法收斂(圖 4.46, 4.49, 4.50, 4.52, 4.53, 4.54)。所剩三題中(圖 4.47, 4.48, 4.51)，也只有 CGLM Method 及 LM Method 兩種方法能收斂而已。

d 部分---由於這部分取的函數較為複雜，所以有一半的題數五種方法皆無法收斂(圖 4.57 , 4.58 , 4.61 , 4.62 , 4.67 , 4.68 , 4.71 , 4.72)。最陡坡降法及共軛梯度法在此部分中有收斂一題(圖 4.63 , 4.64)；而 HWO-OWO Method 也有收斂一題(圖 4.59)。令人無法理解的是，在此部分中有兩題，只是差在神經元個數不同而已，在 EX2 中,神經元較少時是 CGLM Method 較快(圖 4.59)；但神經元調高後，卻變成 LM Method 較快(圖 4.60)。而在 EX4 中，則是呈相反的情形(圖 4.63 , 4.64)。

4 實際例題訓練結果

利用複材夾心板來模擬機翼結構，上下蒙皮視為面板，翼樑肋為心材，並且考慮一複雜函數形式之翼形剖面代替傳統平板結構，來完整模擬機翼結構之動態行為。其自然振動頻率可由下列步驟求得(蓋欣聖，2002；Hwu and Gai, 2003)

$$\|\Gamma_2 r^2 + \Gamma_1 r + \Gamma_0 + \omega^2 \mathbf{I}_0\| = 0 \quad (4.1)$$

由上式特徵值問題可解得特徵值 $r_i(\omega)$ 與其對應的特徵向量 \mathbf{V}_i^* ，其中 i 為模態數。而 Γ_2 、 Γ_1 、 Γ_0 為材料常數所組成之矩陣， \mathbf{I}_0 為慣性項。

由所求得之特徵值及特徵向量，利用展開法(Superposition)可將位移寫成

$$\Delta(y) = \mathbf{V}_1^* e^{r_1 y} + \mathbf{V}_2^* e^{r_2 y} + \mathbf{V}_3^* e^{r_3 y} + \mathbf{V}_4^* e^{r_4 y} + \mathbf{V}_5^* e^{r_5 y} + \mathbf{V}_6^* e^{r_6 y} + \mathbf{V}_7^* e^{r_7 y} + \mathbf{V}_8^* e^{r_8 y} \quad (4.2)$$

將上式所求得之位移表示式代入下列之邊界條件式中

$$\Delta(0) = \mathbf{0} \quad (4.3)$$

$$\mathbf{K}_0\Delta(l) + \mathbf{K}_1\Delta'(l) = \mathbf{0} \quad (4.4)$$

則變為以自然頻率 ω 所構成之特徵值問題，藉由求解由八個邊界條件所構成 8×8 矩陣行列式值為零，即可求出機翼結構之自然頻率。其所對應之八個特徵向量即為 $\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_8$ ，代回(4.2)式即可求得各模態。其中 \mathbf{K}_1 、 \mathbf{K}_0 為材料常數所組成之矩陣。利用上面幾個式子，代入不同的材料常數，即可求得機翼結構之自然頻率。

本文即應用倒傳遞網路來取代上面所敘之計算過程。先利用上述之公式，求出幾組資料來當訓練對，訓練好的網路即可取代上述之公式，以後如遇到相同的問題，即可直接輸入材料常數，而網路的輸出即是複材機翼結構之自然頻率(取5個模態)。訓練對資料如表4.7，4.8，4.9所示，每一行的一到四元素為蒙皮徑度矩陣($\mathbf{Q}_{11}, \mathbf{Q}_{12}, \mathbf{Q}_{22}, \mathbf{Q}_{66}$)，第五及第六元素為翼樑剪力模數($\mathbf{G}_1, \mathbf{G}_2$)。但是由於第五元素之值等於第六元素(表4.7)，根據(3.90)式計算，會產生一個第五列與第六列相依的矩陣，如此，就無法解得聯立方程式之解，因此HWO-OWO Method將無法使用。所以將第六個元素刪除，只取一到五個元素做網路訓練。其訓練結果如圖4.73，4.74所示(隱藏層含10個神經元)，不管展弦比等於二或四都只有LM Method收斂。此時

所設定的收斂條件為疊代 5000 次必須收斂到誤差小於 10^{-5} 。訓練好的網路測試如下：

展弦比=2，測試訓練對輸入為

$$\text{第一組}=[100.1564e09 \ 0.626e09 \ 2.504e09 \ 1.25e09 \ 8e09]^T$$

$$\text{第二組}=[100.1564e09 \ 1.252e09 \ 5.008e09 \ 2.5e09 \ 8e09]^T$$

$$\text{第三組}=[200.313e09 \ 1.252e09 \ 5.008e09 \ 2.5e09 \ 4e09]^T$$

網路輸出與目標值比較如表 4.10。展弦比=4，測試訓練對輸入為

$$\text{第一組}=[100.1564e09 \ 0.626e09 \ 2.504e09 \ 1.25e09 \ 8e09]^T$$

$$\text{第二組}=[100.1564e09 \ 1.252e09 \ 5.008e09 \ 2.5e09 \ 8e09]^T$$

$$\text{第三組}=[200.313e09 \ 1.252e09 \ 5.008e09 \ 2.5e09 \ 4e09]^T$$

網路輸出與目標值比較如表 4.11。若將隱藏層神經元提高到 25 個，當展弦比等於 2 時，LM Method 及 CGLM Method 皆能收斂（圖 4.75）；當展弦比等於 4 時，則變成 LM Method 和 HWO-OWO Method 可收斂（圖 4.76）。若將隱藏層神經元提高到 30 個，當展弦比等於 2 時，依舊是 LM Method 及 CGLM Method 能收斂（圖 4.77）；當展弦比等於 4 時，則變成 LM Method、CGLM Method 和 HWO-OWO Method 皆可收斂（圖 4.78）。

4.5 結果討論

五種方法收斂性如下：

	類型一	類型二	類型三	實例
最陡坡降法	雖能收斂，但疊代次數過多。	掉入局部最小值，無法收斂到所設定之誤差內。	所有例題皆掉入局部最小值。	接近收斂的誤差範圍，但仍無法收斂。
共軛梯度法	能改善最陡坡降法收斂緩慢的缺點。	曲線震盪，無收斂趨勢。	反覆震盪，直至時間終了。	反覆震盪，直至結束。
LM Method	雖能收斂，但所花費時間多於最陡坡降法。	所有例題皆能快速收斂。	在所設定的條件下，約有三分之二例題收斂。	唯一全部皆能收斂的方法。
HOW-OWO Method	快速收斂，此類型最佳選擇。	收斂題數較少，曲線震盪極大。	有時震盪，有時掉入局部最小值，大多數例題無法收斂。	只有神經元 >25 ，展弦比 $=4$ 時，才會收斂。
CGLM Method	雖能收斂，但所花費時間多於最陡坡降法。	約有三分之二的例題，收斂優於 LM Method	情況與 LM Method 差不多，少數例題只停留在共軛梯度法。	神經元較少時，無法收斂。若將神經元提高則能轉為收斂。

其他觀察結果：

1. 在相同的訓練對資料下，OWO-HWO Method 比 LM Method 需要更多的神經元才能完整的凝合曲線。
2. 在共軛梯度法能很快的收斂到 0.05 以下時，CGLM Method 會比 LM Method 有更快的收斂。因為合併法中的 LM Method 的部分，

起始的 $\alpha = 10^{-5}$, 其他的參數不變, 所以導致合併法中的 LM Method 會比合併法中的 LM Method 更快趨近於牛頓法。

3. 神經元愈多, LM Method 疊代一次所需的時間, 會比其他的方法多。因為本文所使用的網路架構皆不大, 所以看不出此缺點。

4. 共軛梯度法和 OWO-HWO Method 的圖形震盪很大, 似乎是因為沒有使用 $\lambda = \lambda_{opt}$, λ 為步長, 也就是學習速率; λ_{opt} 為最佳步長, 也就是經由最佳化方法求得最佳的步長值。本文為求公平起見, 每種方法都將學習速率(步長)固定為 1。

5. OWO-HWO Method 應用的廣泛性不高, 因為:

- 在實際應用題中, 在每一組輸入向量中, 有兩個元素值皆為相等, 這樣引起 OWO-HWO Method 在解方程式時會產生錯誤, 而其他方法都無此困擾。

- 在類題三中, OWO-HWO Method 皆無法在指定的神經元數目下達到收斂, 似乎訓練對較多時, 便無法收斂。

6. 神經元個數不可增加過多, 過多的神經元只會浪費時間, 無法有效的降低誤差。本文舉以最陡坡降法訓練類型二 EX2 為例子, 當隱藏層神經元個數為 40 時, 其誤差最小值為 3.3747×10^{-4} ; 當隱藏層神經元個數為 60 時, 其誤差最小值為 5.6482×10^{-4} ; 但是當隱藏層神經元個數為 100 時, 其誤差最小值卻升高為 0.0052。

第五章

結論與建議

5.1 結論

LM Method 確實擁有不錯的收斂性，但是由於其必須計算出 Hessian Matrix，以致每次疊代所需時間總是比別種方法長，在權重個數少的情況下不明顯，但是一旦權重個數多的話，將會明顯拖長網路訓練時間。而 OWO-HWO Method 在實際應用上有著重大缺失，因為當訓練對輸入的資料剛好有兩個元素相同時(如表 4.7 中的 G_1 、 G_2)，此時在計算(3.90)時，矩陣中會有兩列出現相依的情況，如此將無法求出解來，也就無法使用此方法。在多數的情況下，有限度的增加隱藏層的神經元個數，可以達到更小的誤差值，但是，若增加過多的神經元，不但無法有效的降低誤差，反而會增加網路訓練的時間。

5.2 建議

由於本文中的步長(學習速率)固定為 1，所以有些方法可能需要將最佳化步長求出，才能使其方法發揮較佳的功用，若能將最佳步長一併考慮的話，所得到結果應該可以更讓人信服。此外，本文中的轉換函數固定為

$$f(x) = \frac{1}{1 + e^{-x}}$$

而轉換函數還有其他種類，若能將不同的轉換函數都考慮進去的話，其結果較為完整。



參考文獻

Charalambous, C., “ Conjugate Gradient Algorithm for Efficient Training of Artificial Neural Networks ” , *IEEE Proceedings G*, Vol. 139, No. 3, pp. 301-310, June 1992.

Chen, H. H. , Manry, M. T. and Chandrasekaran, H., “ A Neural Network Training Algorithm utilizing Multiple sets of Linear Equations ” , *Neurocomputing*, Vol. 25, pp. 55-72, 1999.

Chen, M. S. and Liao, F. H., “ Neural Networks Training Using Genetic Algorithms ” , *IEEE International Conference*, Vol. 3 ,pp. 2436-2441, 1998.

Cho, S. Y. and Chow, T. W. S., “ Training Multilayer Neural Networks using Fast Global Learning Algorithm – Least-Squares and Penalized Optimization Methods ” , *Neurocomputing*, Vol. 25 , pp. 115-131, 1999.

Fun, M. H. and Hagan, M. T., “ Levenberg-Marquardt Training for Modular Networks ” , *IEEE International Conference*, Vol. 1 , pp. 468-473, 1996.

Hagan, M. T. and Menhaj, M. B., “ Training Feedforward Networks with the Marquardt Algorithm ” , *IEEE Transaction on Neural Networks*, Vol.5 , No 6. November 1994.

Haykin, S., *Neural Networks A Comprehensive Foundation*, Prentice Hall , 1999.

Hwu, C. and Gai, H.S., “ Vibration Control of Stiffened Composite Wing Structure ” , under preparation.

Liang, Y. C., Feng, D. P., Lee, H. P., Lim, S. P. and Lee, K. H., “ Successive Approximation Training Algorithm for Feedforward Neural

Networks ” , *Neurocomputing*, Vol. 42 , pp. 311-322, 2002.

Rao, S. S., *Engineering Optimization Theory and Practice*Wiley-Interscience, 1996.

Scalero, R. S. and Tepedelenlioglu, N., “ A Fast New Algorithm for Training Feedforward Neural Networks ” , *IEEE Transaction on Signal Processing*, Vol. 40, No. 1, pp.202-210, January 1992.

Towsey, M., Alpsan, D. and Sztriha, L., “ Training a Neural Network With Conjugate Gradient Methods ” , *IEEE Proc. Neural Networks*, Vol. 1, pp. 373-378, 1995.

Yam, Y. F. and Chow, T. W. S., “ Extended backpropagation algorithm ” , *Electronics Letters*, Vol.29, No. 19, pp. 1701-1702, September 1993.

沈宗澤 , “ 以遺傳演算法及共軛梯度法訓練類神經網路 ” , 國立成功大學航空太空工程研究所 , 2001

施柏屹 , “ 倒傳遞類神經網路學習收斂之初步探討 ” , 國立中央大學機械工程研究所 , 2000

張吉甫 , “ 類神經網路在寬頻網路管理上之應用 ” , 元智大學資訊管理研究所 , 2002

葉宜成 , “ 類神經網路模式應用與實作 ” , 儒林 , 1993

蓋欣聖 , “ 複材機翼結構之振動分析及控制 ” , 國立成功大學航空太空工程研究所 , 2002

羅華強 , “ 類神經網路-MATLAB的應用 ” , 清蔚科技 , 2001

附 錄

$$\frac{\partial \delta_2}{\partial w_3} = \frac{\partial}{\partial w_3} [f'(g_2)w_3\delta_3] \quad (\text{A.1})$$

$$= f'(g_2) \left[\delta_3 + \frac{\partial \delta_3}{\partial w_3} w_3 \right] \quad (\text{A.2})$$

式中的 δ_3 對 w_3 的偏微分已經由式子(3.37)求出，所以上式(A.2)已可求得。繼續求

$$\frac{\partial \delta_2}{\partial w_2} = \frac{\partial}{\partial w_2} [f'(g_2)w_3\delta_3] \quad (\text{A.3})$$

$$= w_3 \left[\frac{\partial f'(g_2)}{\partial w_2} \delta_3 + \frac{\partial \delta_3}{\partial w_2} f'(g_2) \right] \quad (\text{A.4})$$

式中

$$\frac{\partial f'(g_2)}{\partial w_2} = \frac{\partial f'(g_2)}{\partial g_2} \frac{\partial g_2}{\partial w_2} = f''(g_2)h_1 \quad (\text{A.5})$$

$$\frac{\partial \delta_3}{\partial w_2} = \frac{\partial}{\partial w_2} [f'(g_3) \cdot (d - h_3)] \quad (\text{A.6})$$

$$= \left[\frac{\partial f'(g_3)}{\partial w_2} (d - h_3) + f'(g_3) \frac{\partial (d - h_3)}{\partial w_2} \right] \quad (\text{A.7})$$

其中

$$\frac{\partial f'(g_3)}{\partial w_2} = \frac{\partial f'(g_3)}{\partial g_3} \frac{\partial g_3}{\partial h_2} \frac{\partial h_2}{\partial g_2} \frac{\partial g_2}{\partial w_2} = f''(g_3)w_3 f'(g_2)h_1 \quad (\text{A.8})$$

$$\frac{\partial (d - h_3)}{\partial w_2} = - \frac{\partial f(g_3)}{\partial g_3} \frac{\partial g_3}{\partial h_2} \frac{\partial h_2}{\partial g_2} \frac{\partial g_2}{\partial w_2} = -f'(g_3)w_3 f'(g_2)h_1 \quad (\text{A.9})$$

所以

$$\frac{\partial \delta_3}{\partial w_2} = \left[-(f'(g_3))^2 + f''(g_3) \cdot (d - h_3) \right] w_3 f'(g_2) h_1 \quad (\text{A.10})$$

將(A.5)及(A.10)代回(A.4)即可求得

$$\frac{\partial \delta_2}{\partial w_2} = w_3 \left\{ f''(g_2) h_1 \delta_3 + \left[-(f'(g_3))^2 + f''(g_3) \cdot (d - h_3) \right] w_3 f'(g_2) h_1 f'(g_2) \right\} \quad (\text{A.11})$$

同理

$$\frac{\partial \delta_1}{\partial w_3} = f'(g_1) w_2 \frac{\partial \delta_2}{\partial w_3} \quad (\text{A.12})$$

$$\frac{\partial \delta_1}{\partial w_2} = f'(g_1) \left[\delta_2 + \frac{\partial \delta_2}{\partial w_2} w_2 \right] \quad (\text{A.13})$$

而最後一項

$$\frac{\partial \delta_1}{\partial w_1} = \left[f'(g_1) w_2 \delta_2 \right] = w_2 \left[\frac{\partial f'(g_1)}{\partial w_1} \delta_2 + \frac{\partial \delta_2}{\partial w_1} f'(g_1) \right] \quad (\text{A.14})$$

式中

$$\frac{\partial f'(g_1)}{\partial w_1} = \frac{\partial f'(g_1)}{\partial g_1} \frac{\partial g_1}{\partial w_1} = f''(g_1) h_0 \quad (\text{A.15})$$

$$\frac{\partial \delta_2}{\partial w_1} = \frac{\partial}{\partial w_1} \left[f'(g_2) w_3 \delta_3 \right] = w_3 \left[\frac{\partial f'(g_2)}{\partial w_1} \delta_3 + \frac{\partial \delta_3}{\partial w_1} f'(g_2) \right] \quad (\text{A.16})$$

其中

$$\frac{\partial f'(g_2)}{\partial w_1} = \frac{\partial f'(g_2)}{\partial g_2} \frac{\partial g_2}{\partial h_1} \frac{\partial h_1}{\partial g_1} \frac{\partial g_1}{\partial w_1} = f''(g_2) w_2 f'(g_1) h_0 \quad (\text{A.17})$$

$$\frac{\partial \delta_3}{\partial w_1} = \frac{\partial}{\partial w_1} \left[f'(g_3) (d - h_3) \right] = \frac{\partial f'(g_3)}{\partial w_1} (d - h_3) + \frac{\partial (d - h_3)}{\partial w_1} f'(g_3) \quad (\text{A.18})$$

其中

$$\begin{aligned}\frac{\partial f'(g_3)}{\partial w_1} &= \frac{\partial f'(g_3)}{\partial g_3} \frac{\partial g_3}{\partial h_2} \frac{\partial h_2}{\partial g_2} \frac{\partial g_2}{\partial h_1} \frac{\partial h_1}{\partial g_1} \frac{\partial g_1}{\partial w_1} \\ &= f''(g_3)w_3 f'(g_2)w_2 f'(g_1)h_0\end{aligned}\quad (\text{A.19})$$

$$\begin{aligned}\frac{\partial(d-h_3)}{\partial w_1} &= -\frac{\partial f(g_3)}{\partial g_3} \frac{\partial g_3}{\partial h_2} \frac{\partial h_2}{\partial g_2} \frac{\partial g_2}{\partial h_1} \frac{\partial h_1}{\partial g_1} \frac{\partial g_1}{\partial w_1} \\ &= -f'(g_3)w_3 f'(g_2)w_2 f'(g_1)h_0\end{aligned}\quad (\text{A.20})$$

所以

$$\frac{\partial \delta_3}{\partial w_1} = \left[f''(g_3)(d-h_3) - (f'(g_3))^2 \right] w_3 f'(g_2)w_2 f'(g_1)h_0 \quad (\text{A.21})$$

$$\begin{aligned}\frac{\partial \delta_2}{\partial w_1} &= w_3 h_0 \left\{ f''(g_2)w_2 f'(g_1) \right. \\ &\quad \left. + \left[f''(g_3)(d-h_3) - (f'(g_3))^2 \right] w_3 (f'(g_2))^2 w_2 f'(g_1) \right\}\end{aligned}\quad (\text{A.22})$$

將(A.15)及(A.22)代回(A.14)即可求得 $\frac{\partial \delta_1}{\partial w_1}$ 。

表 4.1 類型一 基本測試題

類題一		
	輸入	目標值
EX1	$\mathbf{p} = \begin{bmatrix} -1 & -1 & 0 & 1 \\ 1 & -1 & 0 & 0 \end{bmatrix}$	$\mathbf{d} = [1 \quad 1 \quad 0 \quad 0]$
EX2	$\mathbf{p} = \begin{bmatrix} -1 & -1 & 1 & 1 \\ -1 & 1 & -1 & 1 \end{bmatrix}$	$\mathbf{d} = [0 \quad 1 \quad 1 \quad 0]$
EX3	$\mathbf{p} = \begin{bmatrix} -1 & -1 & 2 & 2 \\ 0 & 5 & 0 & 5 \end{bmatrix}$	$\mathbf{d} = [-1 \quad -1 \quad 1 \quad 1]$

表 4.2 類型二 基本函數題

類題二		x 為輸入, $f(x)$ 為目標值
EX1	$f(x) = x$	$x[0,9]$, 間隔為 1
EX2	$f(x) = x^2$	$x[-5,5]$, 間隔為 1
EX3	$f(x) = x^{-1}$	$x[10^{-6},0.1]$, 間隔為 0.01 ; $x[0.1,4.1]$, 間隔為 1
EX4	$f(x) = \sin(x)$	$x[0,2\pi]$, 間隔為 $\frac{\pi}{6}$
EX5	$f(x) = \cos(x)$	$x[0,2\pi]$, 間隔為 $\frac{\pi}{6}$
EX6	$f(x) = \tan(x)$	$x[0,1.25]$, 間隔為 0.25 ; $x[1.3,1.4]$, 間隔為 0.05 ; $x[1.41,1.57]$, 間隔為 0.02
EX7	$f(x) = \log(x)$	$x[10^{-6},0.1]$, 間隔為 0.01 ; $x[0.1,0.5]$, 間隔為 0.1 ; $x[1,3]$, 間隔為 0.5 ;
EX8	$f(x) = e^x$	$x[-5,5]$, 間隔為 1
EX9	$f(x) = \sinh(x)$	$x[0,2\pi]$, 間隔為 $\frac{\pi}{6}$

表 4.2(續)

EX10	$f(x) = \cosh(x)$	$x[0, 2\pi]$, 間隔為 $\frac{\pi}{6}$
EX11	$f(x) = \tanh(x)$	$x[0, 2\pi]$, 間隔為 $\frac{\pi}{6}$
EX12	$f(x) = \sin^{-1}(x)$	$x[-1, 1]$, 間隔為 0.1
EX13	$f(x) = \cos^{-1}(x)$	$x[-1, 1]$, 間隔為 0.1
EX14	$f(x) = \tan^{-1}(x)$	$x[-5, 5]$, 間隔為 0.5

表 4.3 類型三(a) 二對一題型

類題三(a)		x, y 為輸入, $f(x, y)$ 為目標值
EX1	$f(x, y) = x^2 \sin(y)$	$x[-2, 2]$, 間隔為 1 $y[0, 2\pi]$, 間隔為 $\frac{\pi}{6}$
EX2	$f(x, y) = x^2 \tan(y)$	$x[-2, 2]$, 間隔為 1 y 範圍等於類題二 EX6 之 x 範圍
EX3	$f(x, y) = x^2 e^y$	$x[-2, 2]$, 間隔為 1 $y[-5, 5]$, 間隔為 1
EX4	$f(x, y) = x^2 \sinh(y)$	$x[-2, 2]$, 間隔為 1 $y[0, 2\pi]$, 間隔為 $\frac{\pi}{6}$
EX5	$f(x, y) = x^2 \sin^{-1}(y)$	$x[-2, 2]$, 間隔為 1 $y[0, 0.9]$, 間隔為 0.1 ; $y[0.91, 0.99]$, 間隔為 0.01
EX6	$f(x, y) = \sin(x) \tan(y)$	$x[0, 2\pi]$, 間隔為 $\frac{\pi}{6}$ y 範圍等於類題二 EX6 之 x 範圍
EX7	$f(x, y) = \sin(x) e^y$	$x[0, 2\pi]$, 間隔為 $\frac{\pi}{6}$ $y[-5, 5]$, 間隔為 1
EX8	$f(x, y) = \sin(x) \sinh(y)$	$x[0, 2\pi]$, 間隔為 $\frac{\pi}{6}$ $y[0, 2\pi]$, 間隔為 $\frac{\pi}{6}$

表 4.3 (續)

EX9	$f(x, y) = \sin(x) \sin^{-1}(y)$	$x[0, 2\pi]$, 間隔為 $\frac{\pi}{6}$ y 範圍等於類題三(a)EX5 之 y 範圍
EX10	$f(x, y) = \tan(x)e^y$	x 範圍等於類題二 EX6 之 x 範圍 $y[-5, 5]$, 間隔為 1
EX11	$f(x, y) = \tan(x) \sinh(y)$	x 範圍等於類題二 EX6 之 x 範圍 $y[0, 2\pi]$, 間隔為 $\frac{\pi}{6}$
EX12	$f(x, y) = e^x \sinh(y)$	$x[-5, 5]$, 間隔為 1 $y[0, 2\pi]$, 間隔為 $\frac{\pi}{6}$

表 4.4 類型三(b) 一對二題型

類題三(b)		x 為輸入, $f_1(x)$ 及 $f_2(x)$ 為目標值
EX1	$f_1(x) = x^2$ $f_2(x) = \sin(x)$	$x[0, 2\pi]$, 間隔為 $\frac{\pi}{6}$
EX2	$f_1(x) = x^2$ $f_2(x) = \tan(x)$	$x[0, 1.25]$, 間隔為 0.25 ; $x[1.3, 1.4]$, 間隔為 0.05 ; $x[1.41, 1.57]$, 間隔為 0.02
EX3	$f_1(x) = e^x$ $f_2(x) = \sinh(x)$	$x[0, 2\pi]$, 間隔為 $\frac{\pi}{6}$
EX4	$f_1(x) = \tan(x)$ $f_2(x) = \sinh(x)$	同 EX2
EX5	$f_1(x) = \sin(x)$ $f_2(x) = e^x$	$x[0, 2\pi]$, 間隔為 $\frac{\pi}{6}$
EX6	$f_1(x) = x^2 \sin(x)$ $f_2(x) = e^x \cos(x)$	$x[0, 2\pi]$, 間隔為 $\frac{\pi}{6}$
EX7	$f_1(x) = x^2 \sin^{-1}(x)$ $f_2(x) = e^x \cos(x)$	$x[0, 0.9]$, 間隔為 0.1 ; $x[0.91, 0.99]$, 間隔為 0.01
EX8	$f_1(x) = \cos(x) \sin^{-1}(x)$ $f_2(x) = e^x x^2$	$x[0, 0.9]$, 間隔為 0.1 ; $x[0.91, 0.99]$, 間隔為 0.01

表 4.4(續)

EX9	$f_1(x) = \cos(e^x)$ $f_2(x) = \tan(x)$	同 EX2
EX10	$f_1(x) = \tan(e^x)$ $f_2(x) = \cos(x)$	$x[-5, 0.45]$, 間隔為 0.1

表 4.5 類型三(c) 二對二題型

類題三(c)		x, y 輸入, $f_1(x, y)$ 及 $f_2(x, y)$ 為目標值
EX1	$f_1(x, y) = x^2 \sin(y)$ $f_2(x, y) = \tan(x)e^y$	$x[0, 1.25]$, 間隔為 0.25 ; $x[1.3, 1.4]$, 間隔為 0.05 ; $x[1.41, 1.57]$, 間隔為 0.02 $y[0, 2\pi]$, 間隔為 $\frac{\pi}{6}$
EX2	$f_1(x, y) = \sin(x)e^y$ $f_2(x, y) = \tan(x)\sinh(y)$	x 同 EX1 $y[0, 2\pi]$, 間隔為 $\frac{\pi}{6}$
EX3	$f_1(x, y) = e^x \sinh(y)$ $f_2(x, y) = x^2 \sin^{-1}(y)$	$x[-5, 5]$, 間隔為 1 $y[0, 0.9]$, 間隔為 0.1 ; $y[0.91, 0.99]$, 間隔為 0.01
EX4	$f_1(x, y) = \sin(x)\sin(y)$ $f_2(x, y) = e^x e^y$	$x[0, 2\pi]$, 間隔為 $\frac{\pi}{6}$ $y[0, 2\pi]$, 間隔為 $\frac{\pi}{6}$
EX5	$f_1(x, y) = \sinh(x)\sin^{-1}(y)$ $f_2(x, y) = x^2 e^y$	$x[0, 2\pi]$, 間隔為 $\frac{\pi}{6}$ $y[0, 0.9]$, 間隔為 0.1 ; $y[0.91, 0.99]$, 間隔為 0.01
EX6	$f_1(x, y) = \sin(x)\cos(y)$ $f_2(x, y) = e^{xy}$	$x[0, 2\pi]$, 間隔為 $\frac{\pi}{6}$ $y[0, 2\pi]$, 間隔為 $\frac{\pi}{6}$

表 4.5(續)

EX7	$f_1(x, y) = \sin(x)\cos(y)$ $f_2(x, y) = e^{xy}$	$x[0, 2\pi]$, 間隔為 $\frac{\pi}{6}$ $y[-5, 5]$, 間隔為 1
EX8	$f_1(x, y) = \tan(x)y$ $f_2(x, y) = e^x \cos(y)$	x 同 EX1 $y[0, 2\pi]$, 間隔為 $\frac{\pi}{6}$

表 4.6 類型三(d) 二對一混合題型

類題三(d)		x, y 為輸入, $f(x, y)$ 為目標值
EX1	$f(x, y) = \sin(xy)$	$x, y[0, 2\pi]$, 間隔為 $\frac{\pi}{6}$
EX2	$f(x, y) = e^{-xy}$	$x, y[-5, 5]$, 間隔為 1
EX3	$f(x, y) = \tan(xy)$	$x[0, 1.25]$, 間隔為 0.25 ; $x[1.3, 1.4]$, 間隔為 0.05 ; $x[1.41, 1.57]$, 間隔為 0.02 $y = x$
EX4	$f(x, y) = \sinh(xy)$	$x, y[0, 2\pi]$, 間隔為 $\frac{\pi}{6}$
EX5	$f(x, y) = \sin^{-1}(xy)$	$x[0, 0.9]$, 間隔為 0.1 ; $x[0.91, 0.99]$, 間隔為 0.01 $y = x$
EX6	$f(x, y) = \cos(xy)$	$x, y[0, 2\pi]$, 間隔為 $\frac{\pi}{6}$
EX7	$f(x, y) = \sin(x \cdot \cos(y))$	$x, y[0, 2\pi]$, 間隔為 $\frac{\pi}{6}$
EX8	$f(x, y) = \sin(e^{xy})$	$x, y[-5, 5]$, 間隔為 1

表 4.7 訓練對輸入資料

訓練對資料(輸入)	第 1 組	第 2 組	第 3 組	第 4 組	第 5 組	第 6 組	第 7 組	第 8 組	第 9 組	第 10 組	第 11 組
蒙皮勁度矩陣 Q11	2.00E+11	1.60E+11	1.20E+11	8.01E+10	4.01E+10	1.60E+11	1.20E+11	8.01E+10	4.01E+10	2.00E+11	2.00E+11
蒙皮勁度矩陣 Q12	1.25E+09	1.00E+09	7.50E+08	5.01E+08	2.50E+08	1.25E+09	1.25E+09	1.25E+09	1.25E+09	1.13E+09	6.30E+08
蒙皮勁度矩陣 Q22	5.01E+09	4.01E+09	3.00E+09	2.00E+09	1.00E+09	5.01E+09	5.01E+09	5.01E+09	5.01E+09	5.01E+09	5.01E+09
蒙皮勁度矩陣 Q66	2.50E+09	2.00E+09	1.50E+09	1.00E+09	5.00E+08	2.50E+09	2.50E+09	2.50E+09	2.50E+09	2.50E+09	2.50E+09
剪力模數 G1	8.00E+09	8.00E+09	8.00E+09	8.00E+09	8.00E+09	8.00E+09	8.00E+09	8.00E+09	8.00E+09	8.00E+09	8.00E+09
剪力模數 G2	8.00E+09	8.00E+09	8.00E+09	8.00E+09	8.00E+09	8.00E+09	8.00E+09	8.00E+09	8.00E+09	8.00E+09	8.00E+09

訓練對資料(輸入)	第 12 組	第 13 組	第 14 組	第 15 組	第 16 組	第 17 組	第 18 組	第 19 組	第 20 組	第 21 組	第 22 組
蒙皮勁度矩陣 Q11	2.00E+11	2.00E+11	2.00E+11	2.00E+11	2.00E+11	2.00E+11	2.00E+11	2.00E+11	2.00E+11	2.00E+11	2.00E+11
蒙皮勁度矩陣 Q12	1.30E+08	1.00E+07	1.13E+05	1.25E+09	1.25E+09	1.25E+09	1.25E+09	1.25E+09	1.25E+09	1.25E+09	1.25E+09
蒙皮勁度矩陣 Q22	5.01E+09	5.01E+09	5.01E+09	4.51E+09	5.00E+08	5.00E+07	5.01E+05	5.01E+09	5.01E+09	5.01E+09	5.01E+09
蒙皮勁度矩陣 Q66	2.50E+09	2.50E+09	2.50E+09	2.50E+09	2.50E+09	2.50E+09	2.50E+09	2.25E+09	2.50E+08	3.00E+07	2.50E+05
剪力模數 G1	8.00E+09	8.00E+09	8.00E+09	8.00E+09	8.00E+09	8.00E+09	8.00E+09	8.00E+09	8.00E+09	8.00E+09	8.00E+09
剪力模數 G2	8.00E+09	8.00E+09	8.00E+09	8.00E+09	8.00E+09	8.00E+09	8.00E+09	8.00E+09	8.00E+09	8.00E+09	8.00E+09

表 4.7 (續)

訓練對資料(輸入)	第 23 組	第 24 組	第 25 組	第 26 組
蒙皮勁度矩陣 Q11	2.00E+11	2.00E+11	2.00E+11	2.00E+11
蒙皮勁度矩陣 Q12	1.25E+09	1.25E+09	1.25E+09	1.25E+09
蒙皮勁度矩陣 Q22	5.01E+09	5.01E+09	5.01E+09	5.01E+09
蒙皮勁度矩陣 Q66	2.50E+09	2.50E+09	2.50E+09	2.50E+09
剪力模數 G1	6.40E+09	4.80E+09	3.20E+09	1.60E+09
剪力模數 G2	6.40E+09	4.80E+09	3.20E+09	1.60E+09

機翼展弦比等於 2 及機翼展弦比等於 4 的訓練對輸入皆為表 4.7

表 4.8 訓練對輸出資料（機翼展弦比等於二）

訓練對資料(輸出)	第 1 組	第 2 組	第 3 組	第 4 組	第 5 組	第 6 組	第 7 組	第 8 組	第 9 組	第 10 組	第 11 組
自然頻率(mode1)	2.31E+02	2.18E+02	2.04E+02	1.87E+02	1.68E+02	2.19E+02	2.05E+02	1.90E+02	1.72E+02	2.31E+02	2.31E+02
自然頻率(mode2)	4.70E+02	4.40E+02	4.02E+02	3.53E+02	2.82E+02	4.42E+02	4.07E+02	3.63E+02	3.02E+02	4.66E+02	4.70E+02
自然頻率(mode3)	9.71E+02	9.43E+02	9.11E+02	8.67E+02	8.10E+02	9.46E+02	9.13E+02	8.75E+02	8.23E+02	9.71E+02	9.71E+02
自然頻率(mode4)	1.45E+03	1.37E+03	1.27E+03	1.15E+03	9.76E+02	1.37E+03	1.28E+03	1.17E+03	1.02E+03	1.44E+03	1.45E+03
自然頻率(mode5)	2.03E+03	1.98E+03	1.96E+03	1.86E+03	1.79E+03	2.01E+03	1.94E+03	1.90E+03	1.81E+03	2.05E+03	2.05E+03

訓練對資料(輸出)	第 12 組	第 13 組	第 14 組	第 15 組	第 16 組	第 17 組	第 18 組	第 19 組	第 20 組	第 21 組	第 22 組
自然頻率(mode1)	2.31E+02	2.30E+02	2.30E+02	2.31E+02	2.29E+02	2.29E+02	2.29E+02	2.31E+02	2.30E+02	2.30E+02	2.30E+02
自然頻率(mode2)	4.68E+02	4.72E+02	4.72E+02	4.70E+02	4.67E+02	4.71E+02	4.67E+02	4.70E+02	4.64E+02	4.64E+02	4.64E+02
自然頻率(mode3)	9.71E+02	9.71E+02	9.71E+02	9.71E+02	9.68E+02	9.69E+02	9.68E+02	9.71E+02	9.70E+02	9.70E+02	9.70E+02
自然頻率(mode4)	1.44E+03	1.45E+03	1.45E+03	1.45E+03	1.44E+03	1.44E+03	1.44E+03	1.44E+03	1.43E+03	1.43E+03	1.43E+03
自然頻率(mode5)	2.05E+03	2.04E+03	2.04E+03	2.03E+03	2.02E+03	2.04E+03	2.03E+03	2.03E+03	2.05E+03	2.05E+03	2.05E+03

表 4.8 (續)

訓練對資料(輸出)	第 23 組	第 24 組	第 25 組	第 26 組
自然頻率(mode1)	2.25E+02	2.16E+02	2.00E+02	1.68E+02
自然頻率(mode2)	4.50E+02	4.22E+02	3.83E+02	3.23E+02
自然頻率(mode3)	9.05E+02	8.22E+02	7.12E+02	5.48E+02
自然頻率(mode4)	1.37E+03	1.28E+03	1.15E+03	9.70E+02
自然頻率(mode5)	1.87E+03	1.66E+03	1.40E+03	1.03E+03

表 4.9 訓練對輸出資料（機翼展弦比等於四）

訓練對資料(輸出)	第 1 組	第 2 組	第 3 組	第 4 組	第 5 組	第 6 組	第 7 組	第 8 組	第 9 組	第 10 組	第 11 組
自然頻率(mode1)	63.0005	58.9608	54.5464	49.6529	44.1076	59.1426	54.9446	50.3195	45.1276	62.9931	62.9625
自然頻率(mode2)	225.7098	208.6061	187.7191	160.8936	122.878	209.6865	190.3897	166.1264	133.2263	225.7912	226.1407
自然頻率(mode3)	336.3947	317.1855	301.5438	279.5764	252.8999	320.9613	303.2534	282.6836	258.046	336.3672	336.2514
自然頻率(mode4)	690.7711	642.7596	584.0482	509.1639	403.6593	645.6386	591.5901	523.4978	431.8106	690.9946	691.9386
自然頻率(mode5)	795.7747	761.6665	728.1963	684.7645	629.1329	763.823	730.625	691.8293	640.9579	790.638	790.4987

訓練對資料(輸出)	第 12 組	第 13 組	第 14 組	第 15 組	第 16 組	第 17 組	第 18 組	第 19 組	第 20 組	第 21 組	第 22 組
自然頻率(mode1)	62.9324	62.9254	62.9244	62.9552	62.5897	62.5484	62.5438	62.9698	62.7236	62.696	62.6926
自然頻率(mode2)	226.4783	226.5586	226.568	225.5217	224.0403	223.868	223.8516	225.3494	222.4443	222.1112	222.0811
自然頻率(mode3)	336.1385	336.1116	336.1077	336.2204	334.7815	334.6226	334.604	336.2761	335.3198	335.2121	335.1987
自然頻率(mode4)	692.8557	693.068	693.0855	690.2363	686.0979	685.6014	685.5538	689.7646	681.6426	680.7129	680.623
自然頻率(mode5)	790.3657	790.3326	790.3286	790.4422	787.7412	787.5366	787.5096	790.5117	789.2706	789.1343	789.1172

表 4.9 (續)

訓練對資料(輸出)	第 23 組	第 24 組	第 25 組	第 26 組
自然頻率(mode1)	62.4436	61.6538	60.1542	56.2035
自然頻率(mode2)	217.9997	207.1302	190.5149	161.0481
自然頻率(mode3)	323.6344	305.7071	277.6698	226.5287
自然頻率(mode4)	663.5968	626.7357	573.037	468.0816
自然頻率(mode5)	745.7364	686.7124	603.1814	483.6561

表 4.10 機翼展弦比等於二之測試結果

	第 1 組測試例	第 1 組測試例	第 2 組測試例	第 2 組測試例	第 3 組測試例	第 3 組測試例
	Marquardt	目標值(程式)	Marquardt	目標值(程式)	Marquardt	目標值(程式)
自然頻率(mode1)	0.1675*e+3	0.1959*e+3	0.188*e+3	0.1977*e+3	0.2037*e+3	0.2092*e+3
自然頻率(mode2)	0.3582*e+3	0.3795*e+3	0.3843*e+3	0.3862*e+3	0.4043*e+3	0.4044*e+3
自然頻率(mode3)	0.8666*e+3	0.891*e+3	0.8805*e+3	0.8969*e+3	0.764*e+3	0.7716*e+3
自然頻率(mode4)	1.0823*e+3	1.2095*e+3	1.1782*e+3	1.2228*e+3	1.17*e+3	1.2211*e+3
自然頻率(mode5)	2.0162*e+3	1.9277*e+3	1.8708*e+3	1.973*e+3	1.5308*e+3	1.5427*e+3

表 4.11 機翼展弦比等於四之測試結果

	第 1 組測試例	第 1 組測試例	第 2 組測試例	第 2 組測試例	第 3 組測試例	第 3 組測試例
	Marquardt	目標值(程式)	Marquardt	目標值(程式)	Marquardt	目標值(程式)
自然頻率(mode1)	54.1589	52.1676	52.2829	52.6929	60.9038	61.0419
自然頻率(mode2)	169.9062	175.2541	171.5351	179.0248	194.0944	199.8049
自然頻率(mode3)	286.0828	291.0453	277.7785	293.4169	281.4398	293.434
自然頻率(mode4)	536.2743	549.1624	539.3075	559.5056	585.845	602.715
自然頻率(mode5)	696.1816	708.1041	690.8283	713.4103	629.9977	649.0155

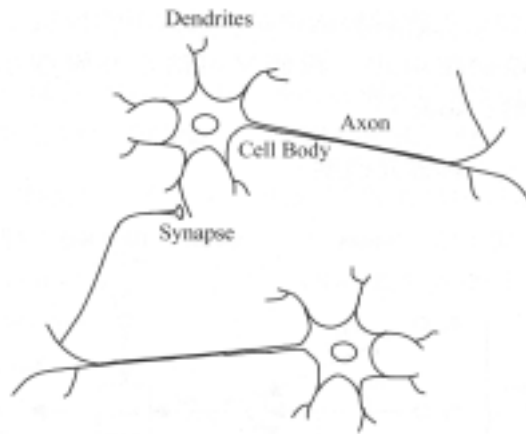


圖 2.1 生物神經元(羅華強, 2001)

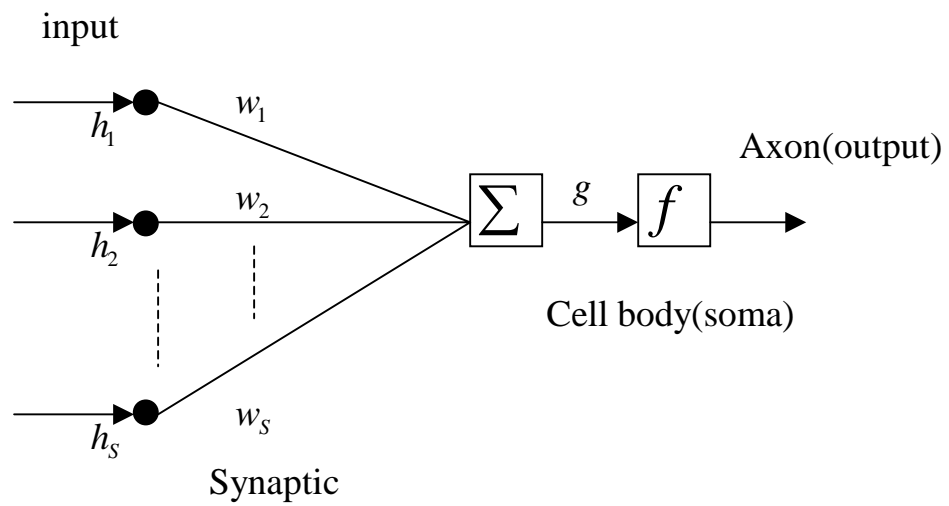


圖 2.2 以人工神經元模擬生物神經元

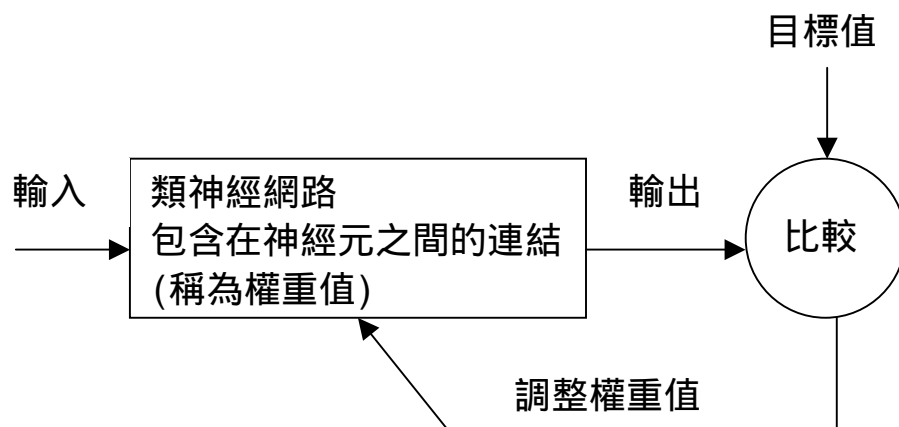
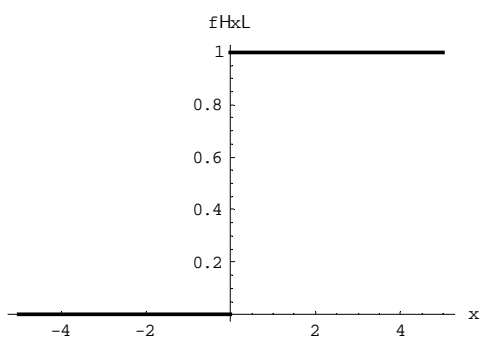
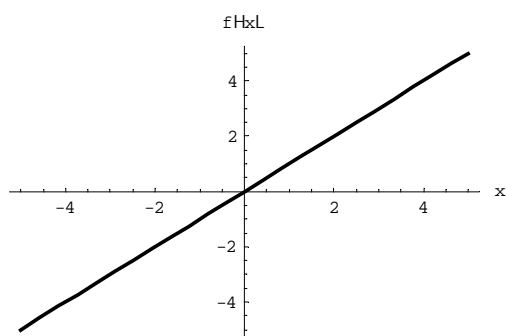


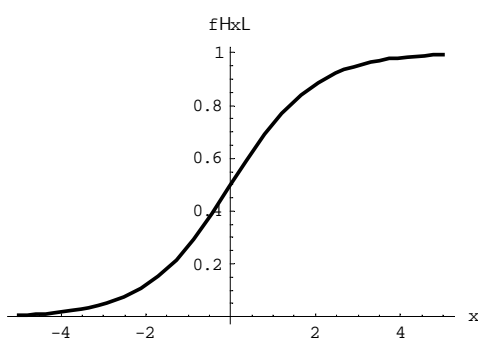
圖 2.3 類神經網路學習程序



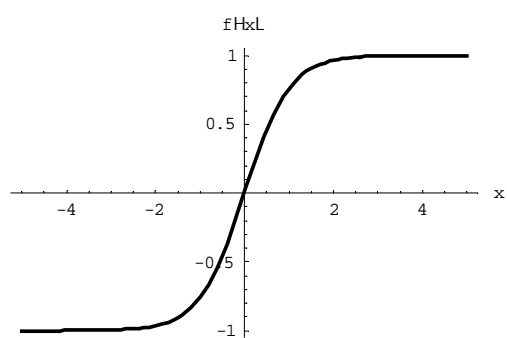
$$f(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases}$$



$$f(x) = x$$



$$f(x) = \frac{1}{1 + e^{-x}}$$



$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

圖 2.4 轉換函數種類

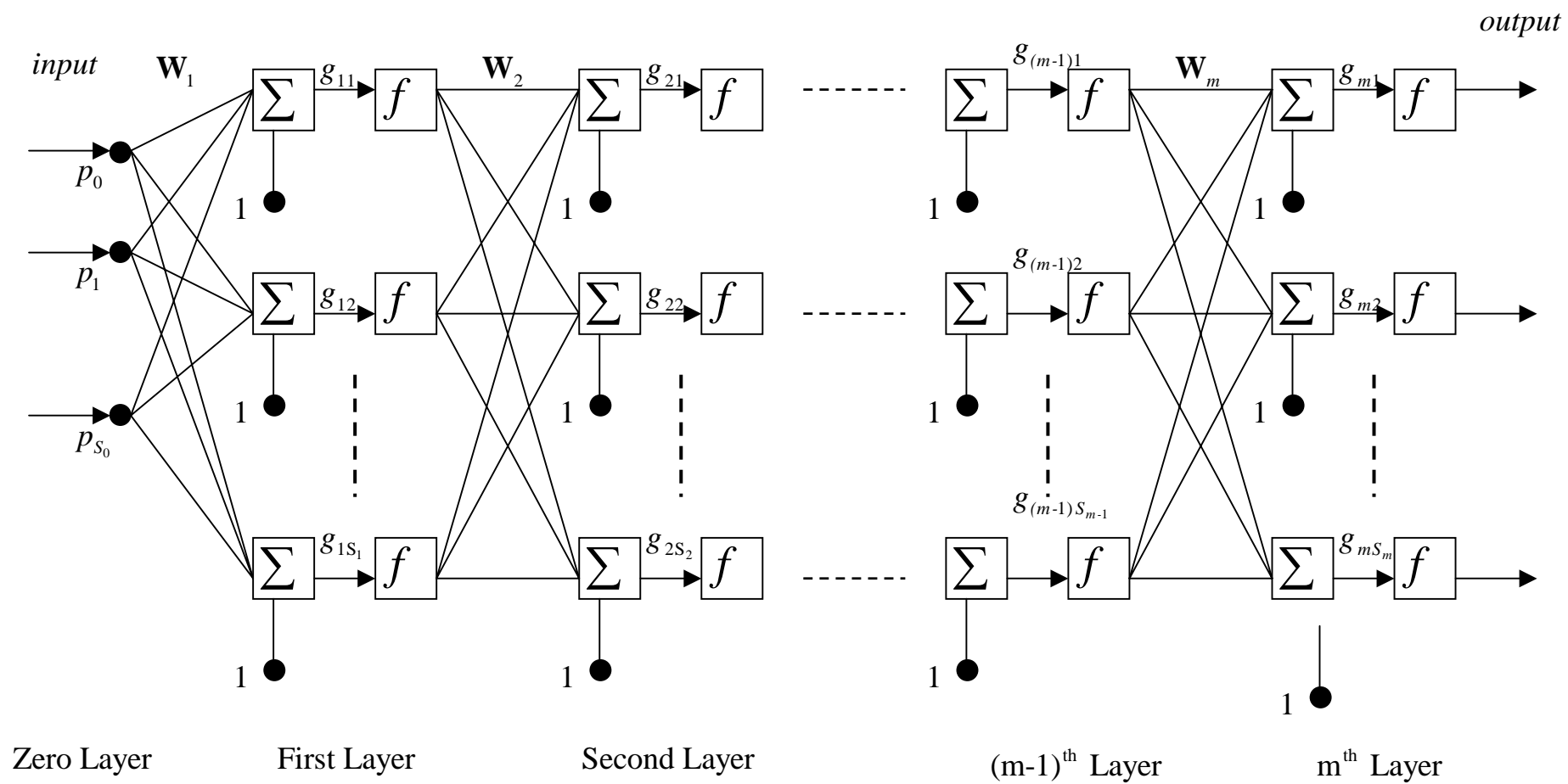


圖 2.5 網路多層架構

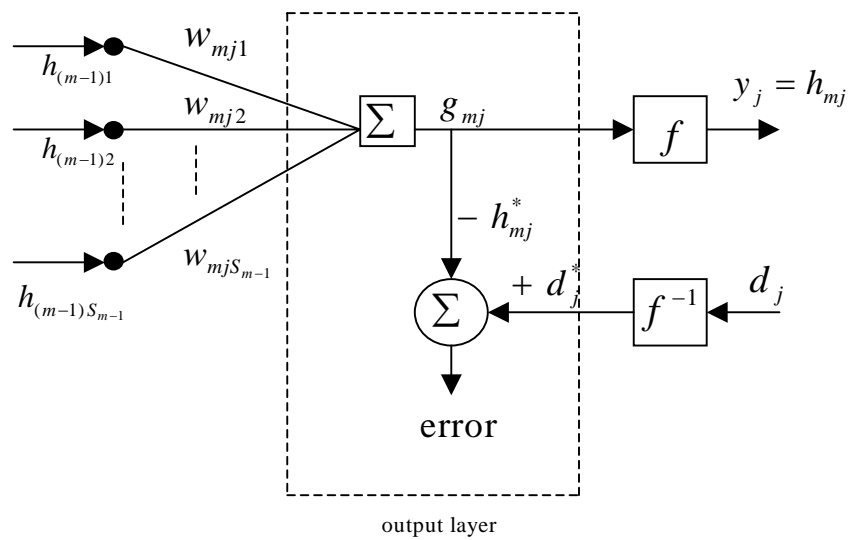


圖 3.1(a) 不含轉換函數之輸出層目標函數

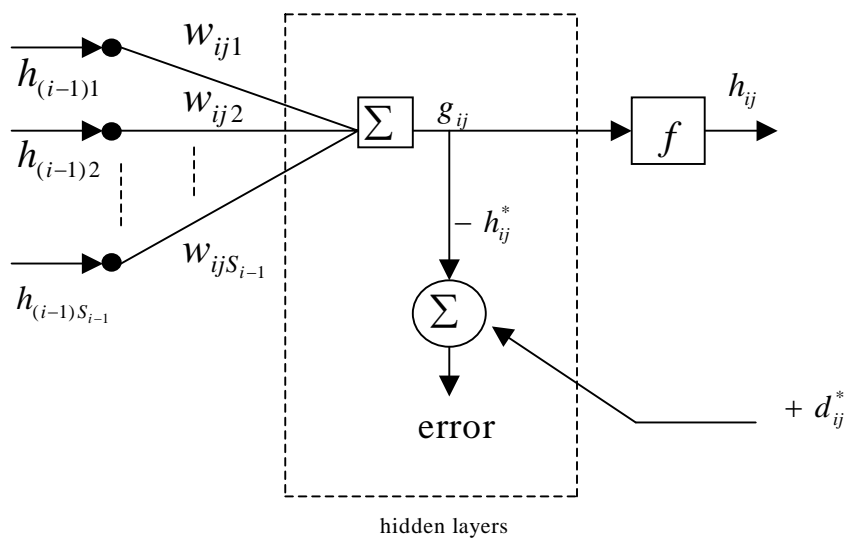


圖 3.1(b) 不含轉換函數之隱藏層目標函數

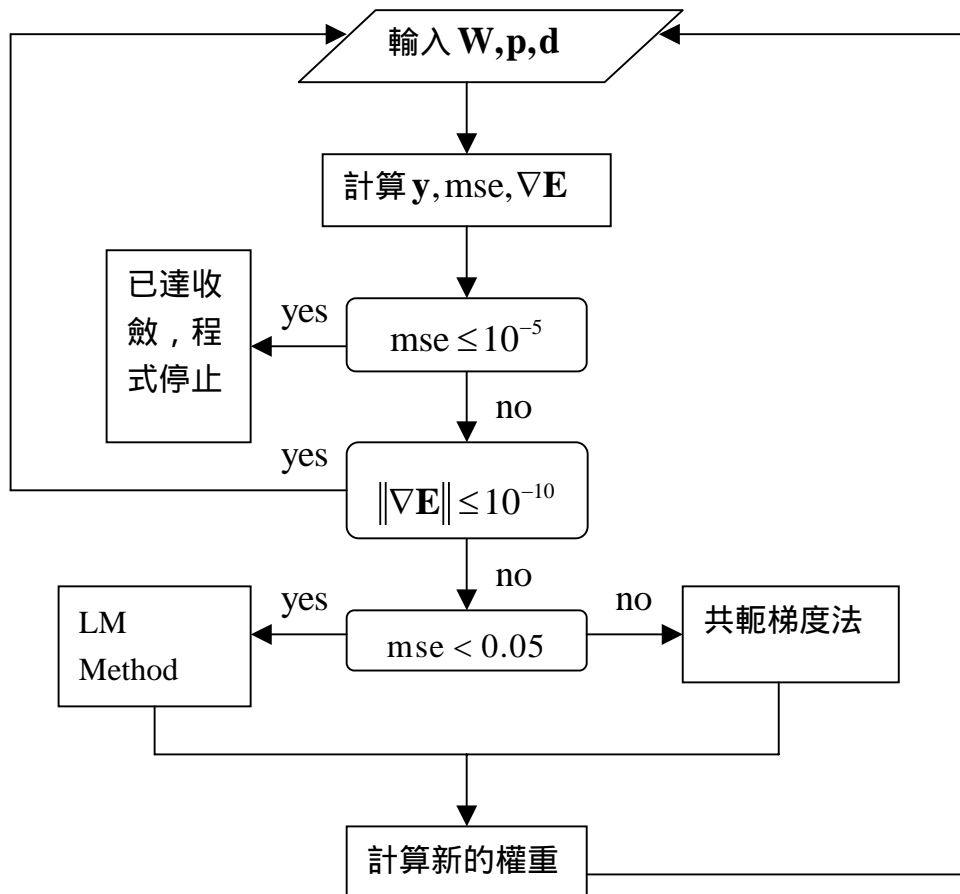


圖 3.2 CGLM Method 流程圖

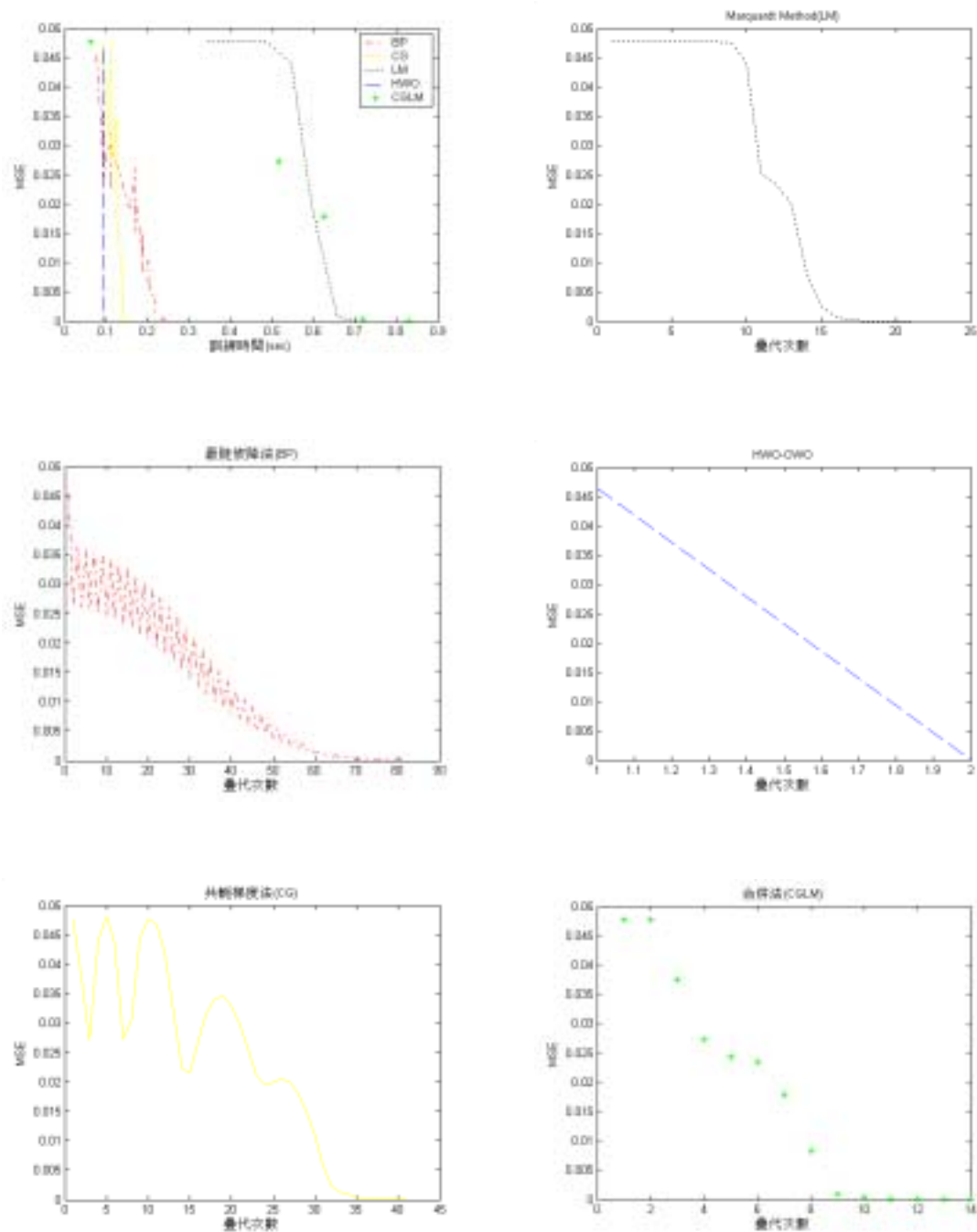


圖 4.1 類型一 EX1 (2-10-1)

其中之(2-10-1)表示網路架構為，輸入層神經元數為 2，隱藏層神經元為 10，輸出層神經元 1。

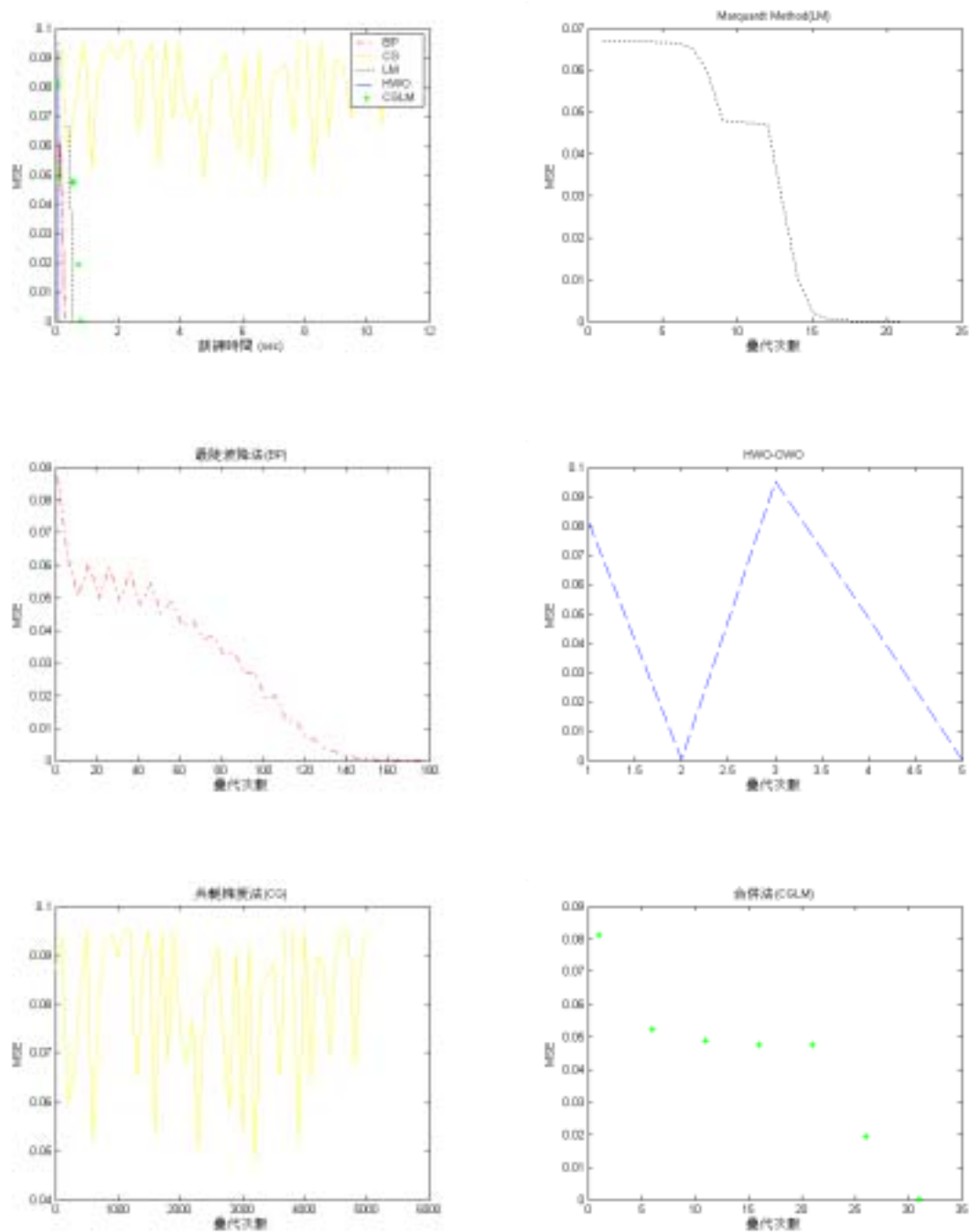


圖 4.2 類型一 EX2 (2-5-1)

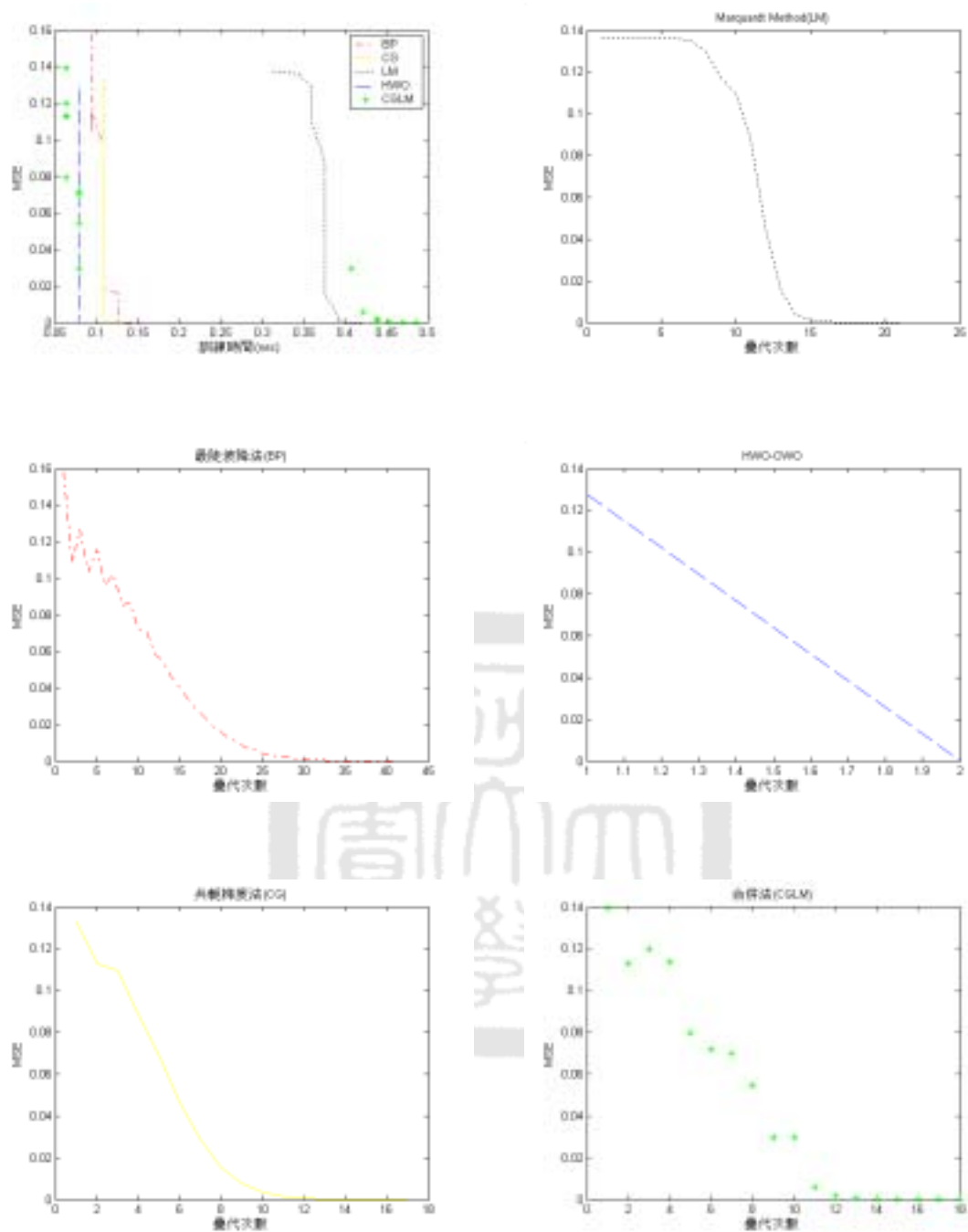


圖 4.3 類型一 EX3 (2-2-1)

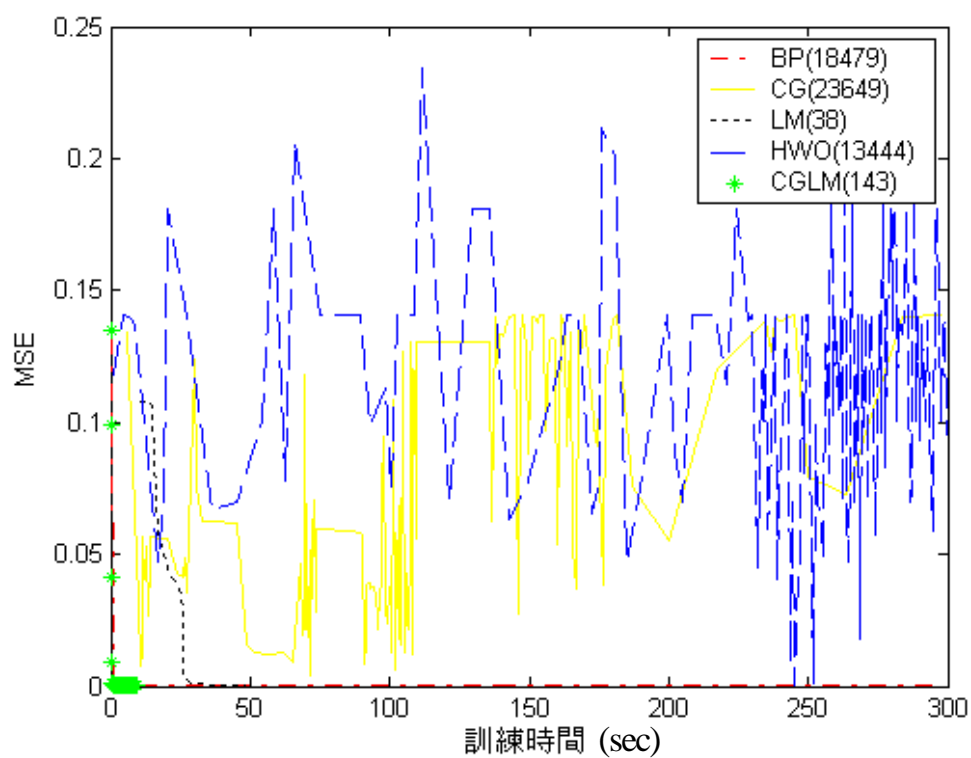


圖 4.4a 類型二 EX1 (1-8-1)

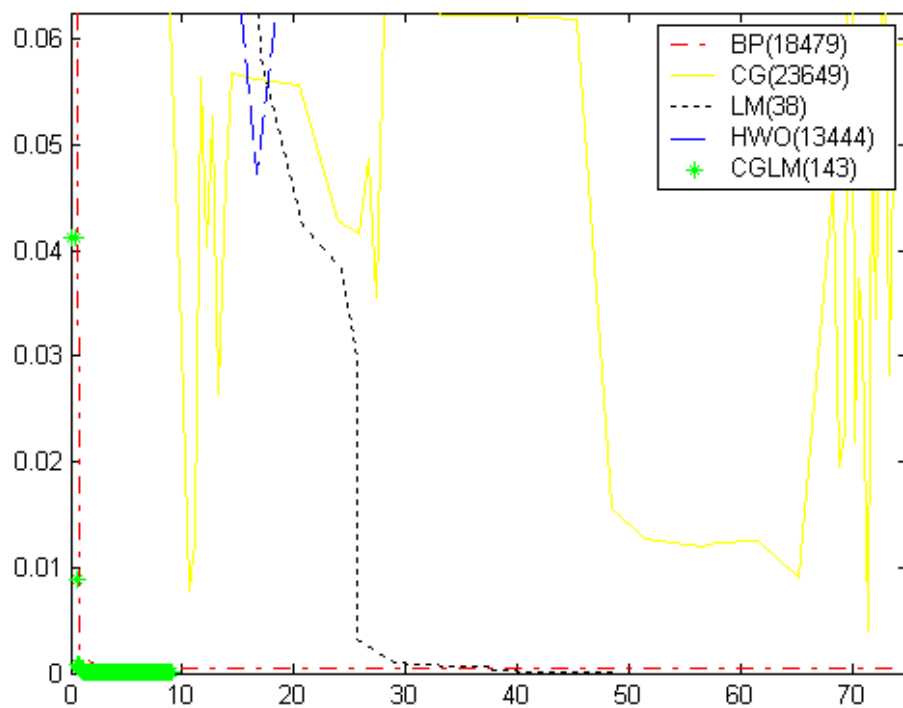


圖 4.4b 放大圖[類型二 EX1 (1-8-1)]

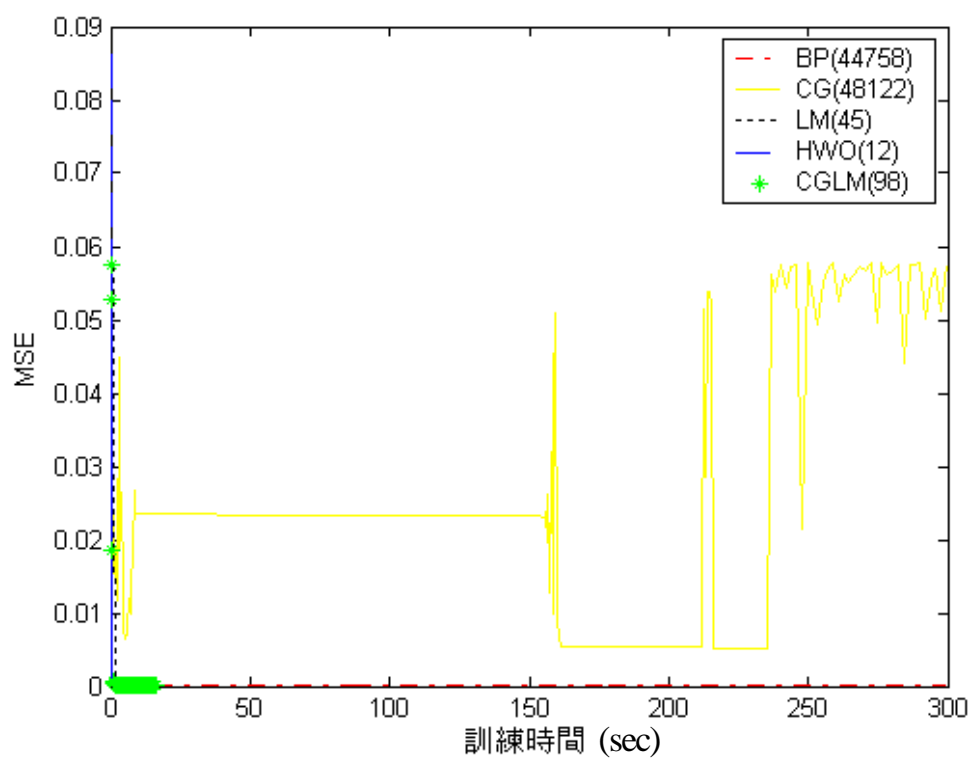


圖 4.5a 類型二 EX1 (1-20-1)

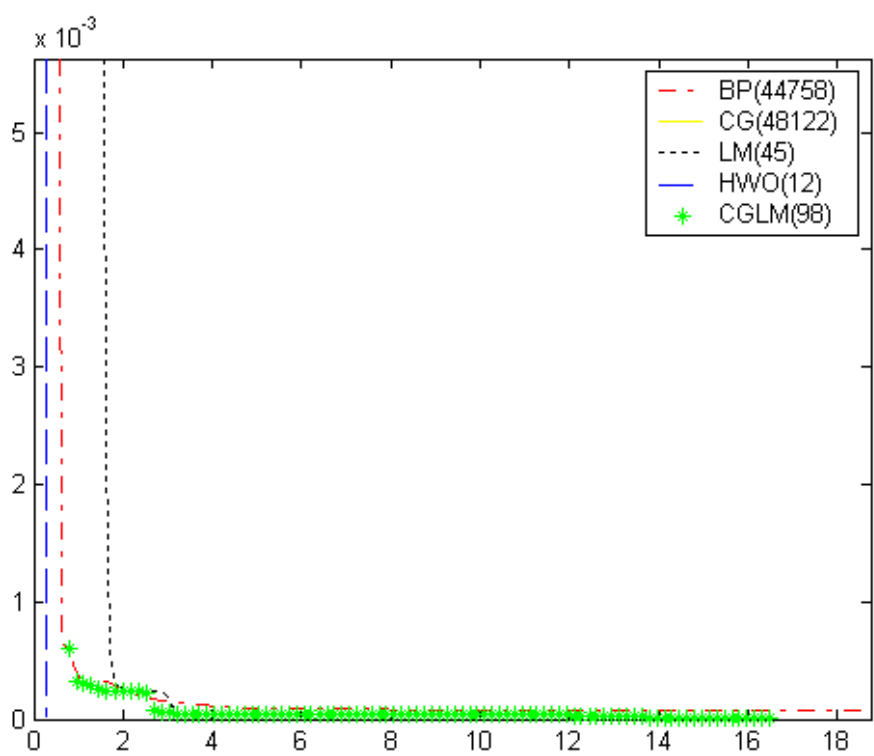


圖 4.5b 放大圖[類型二 EX1 (1-20-1)]

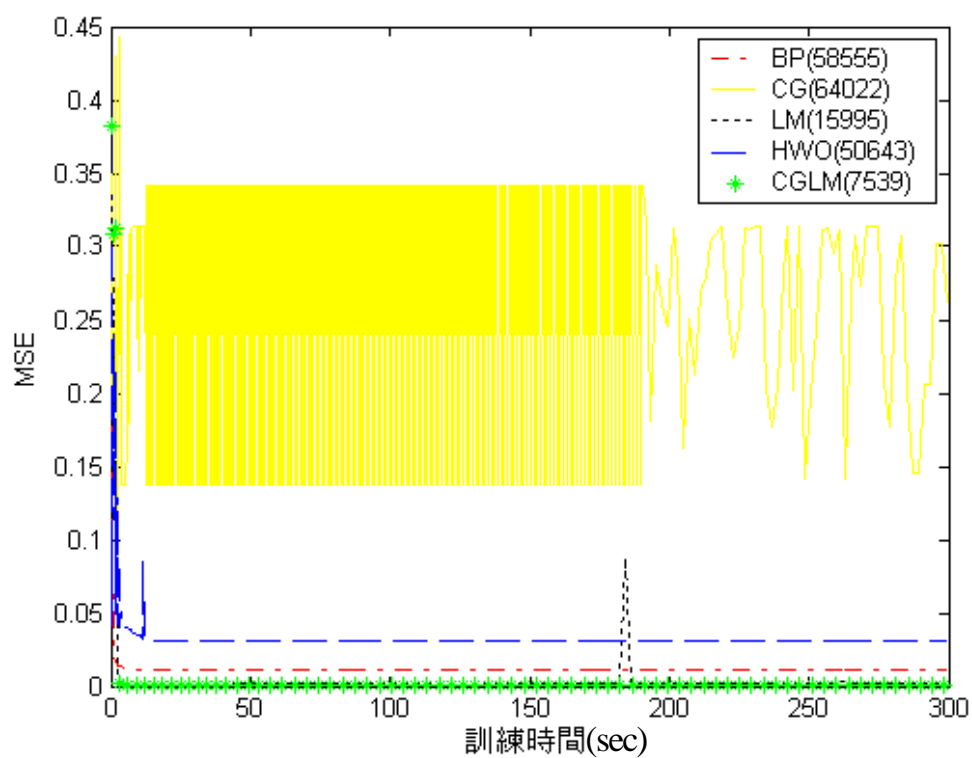


圖 4.6a 類型二 EX2 (1-3-1)

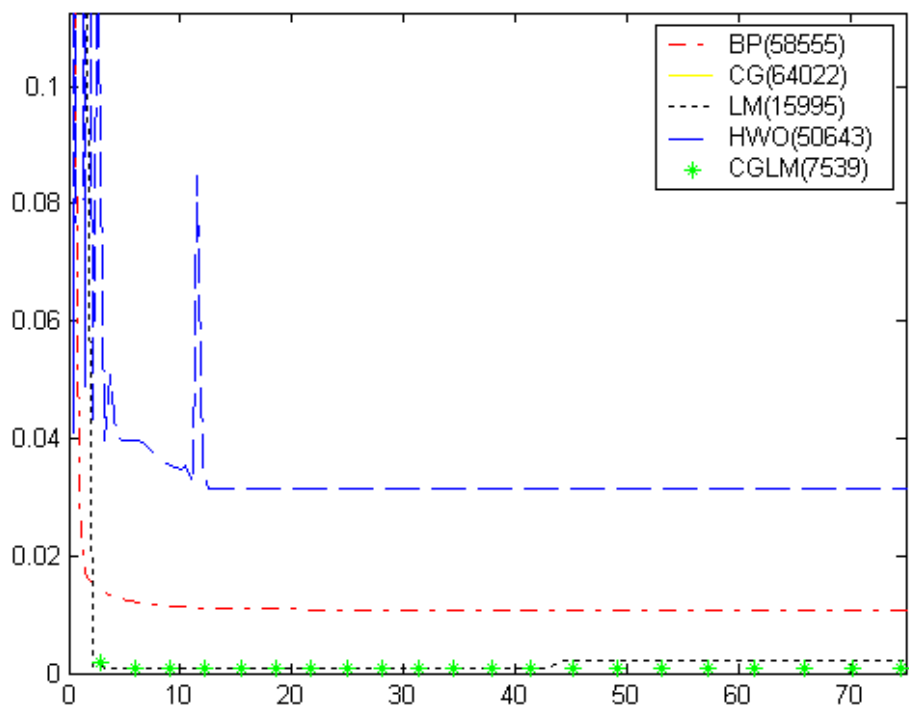


圖 4.6b 放大圖[類型二 EX2 (1-3-1)]

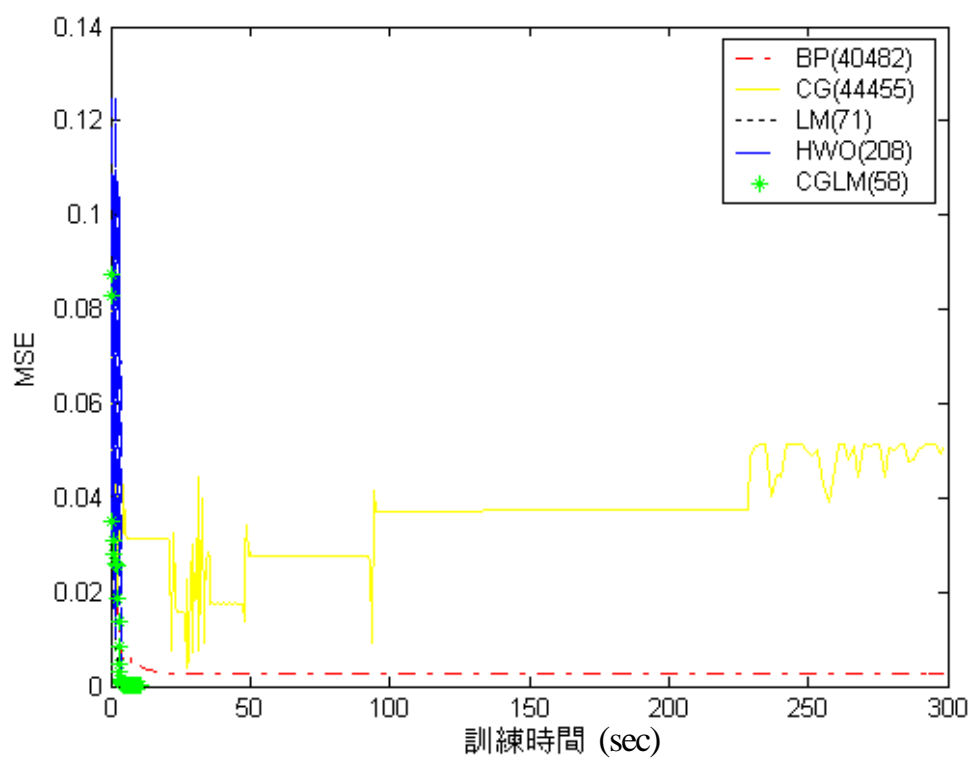


圖 4.7a 類型二 EX2 (1-20-1)

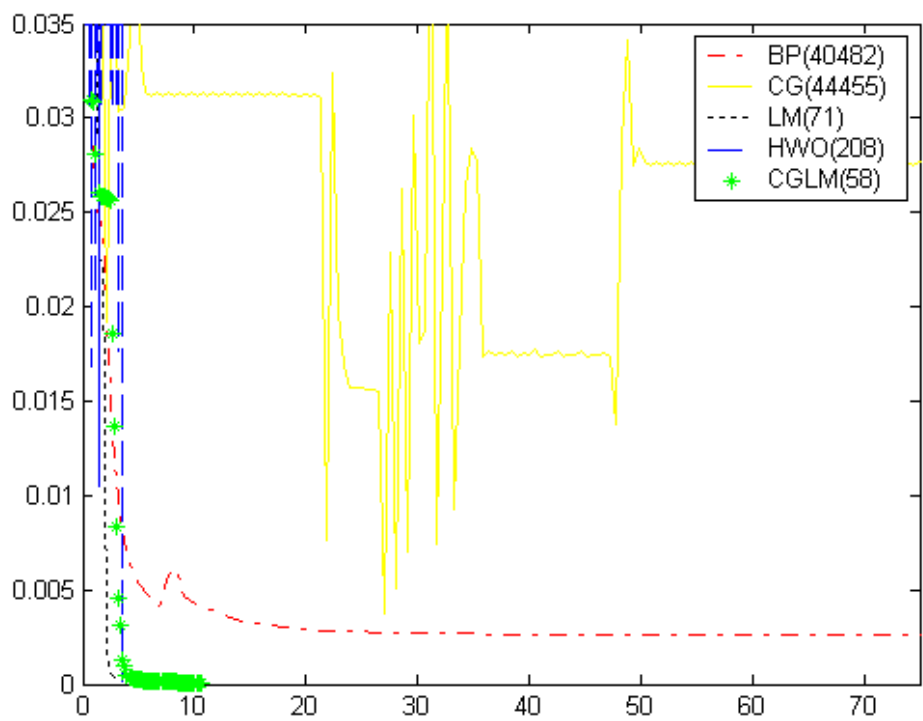


圖 4.7b 放大圖[類型二 EX2 (1-20-1)]

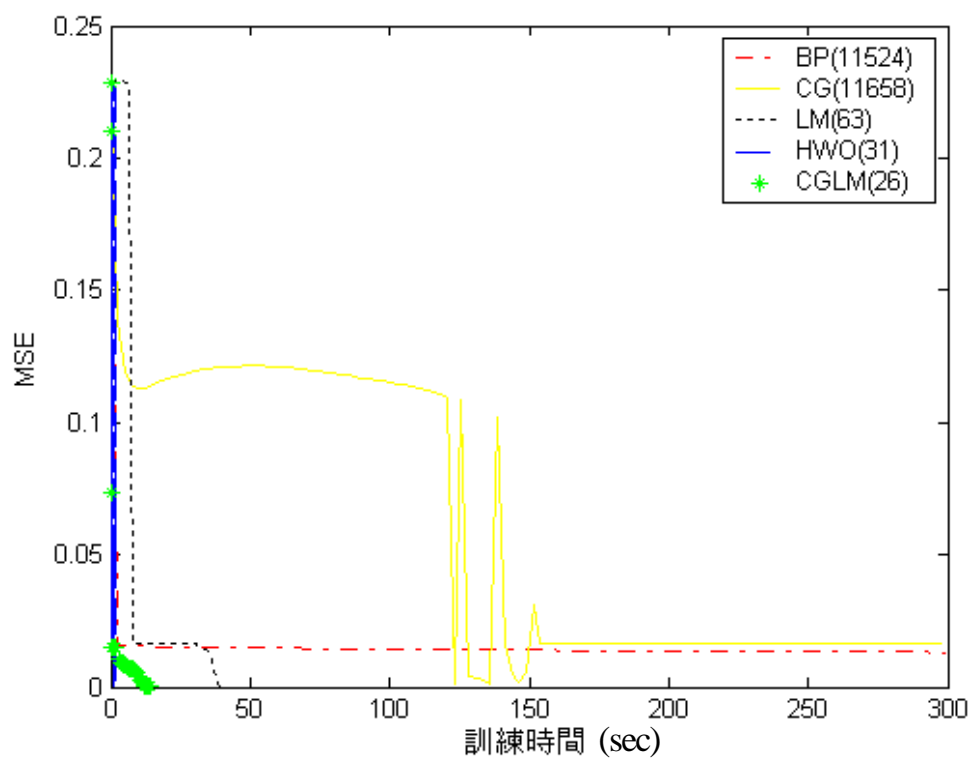


圖 4.8 類型二 EX3 (1-20-1)

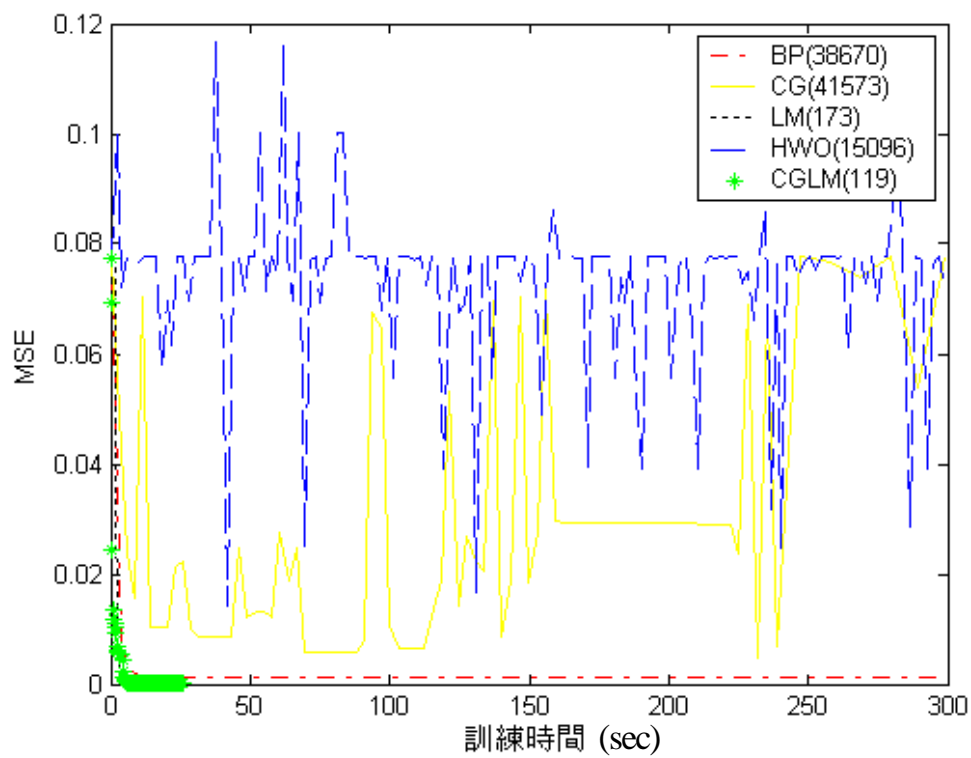


圖 4.9 類型二 EX4 (1-20-1)

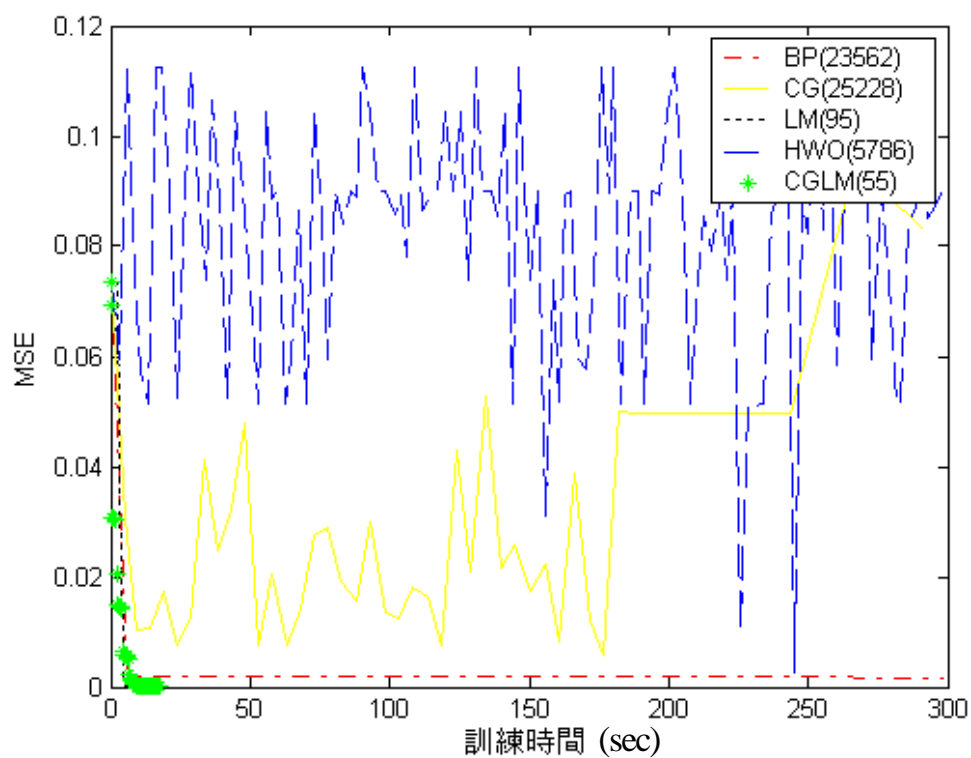


圖 4.10 類型二 EX5 (1-20-1)

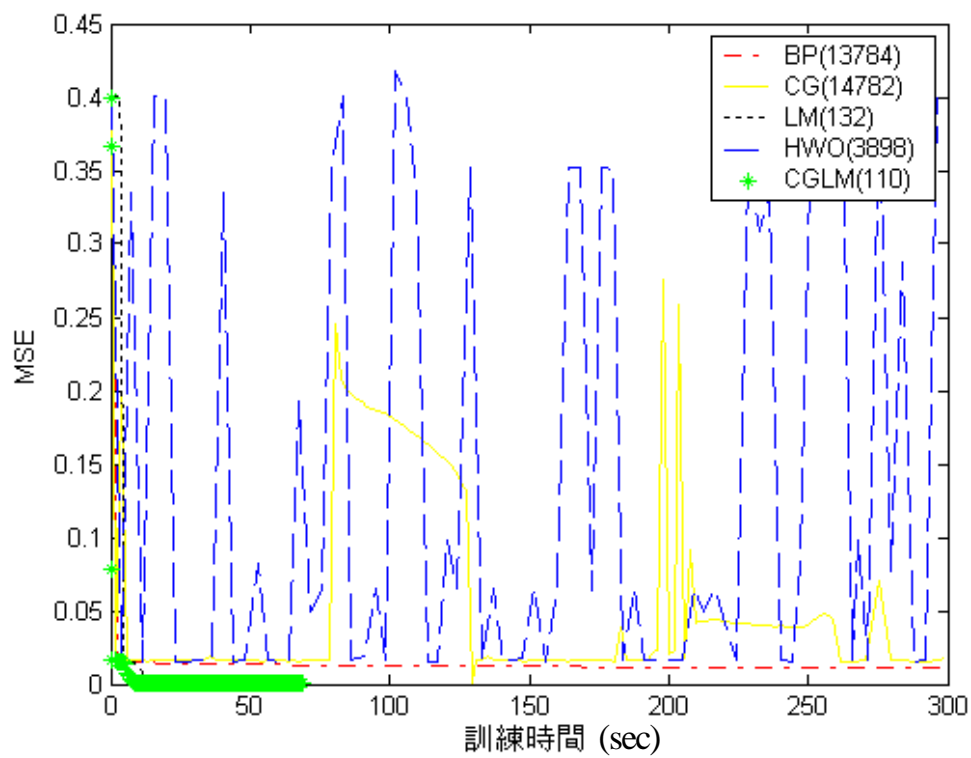


圖 4.11 類型二 EX6 (1-20-1)

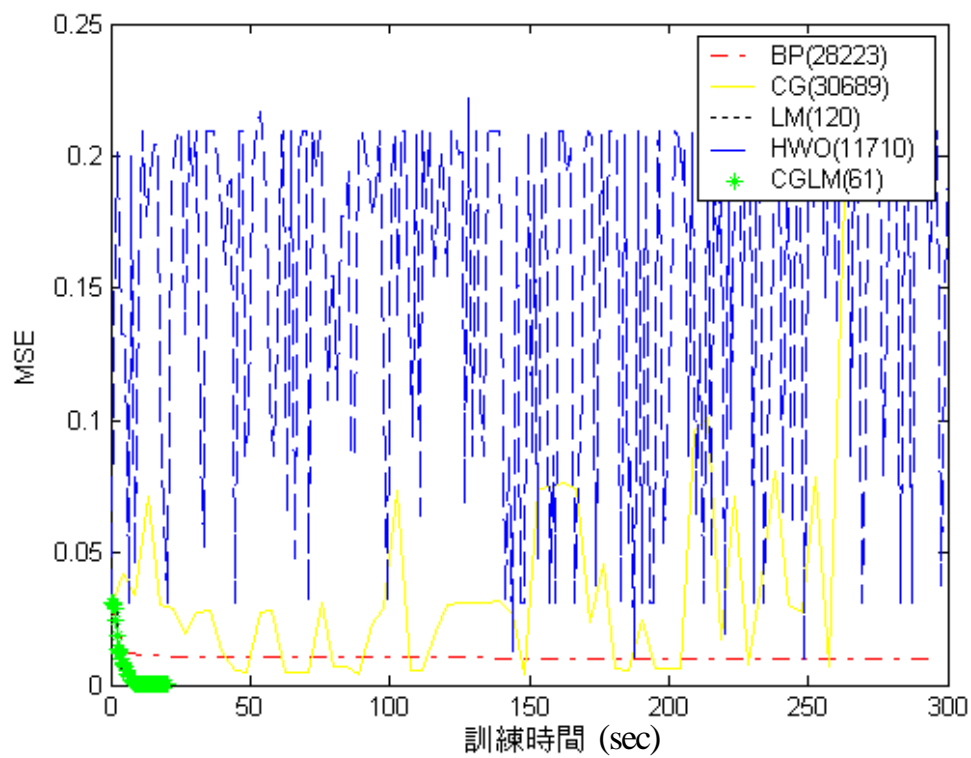


圖 4.12 類型二 EX7 (1-20-1)

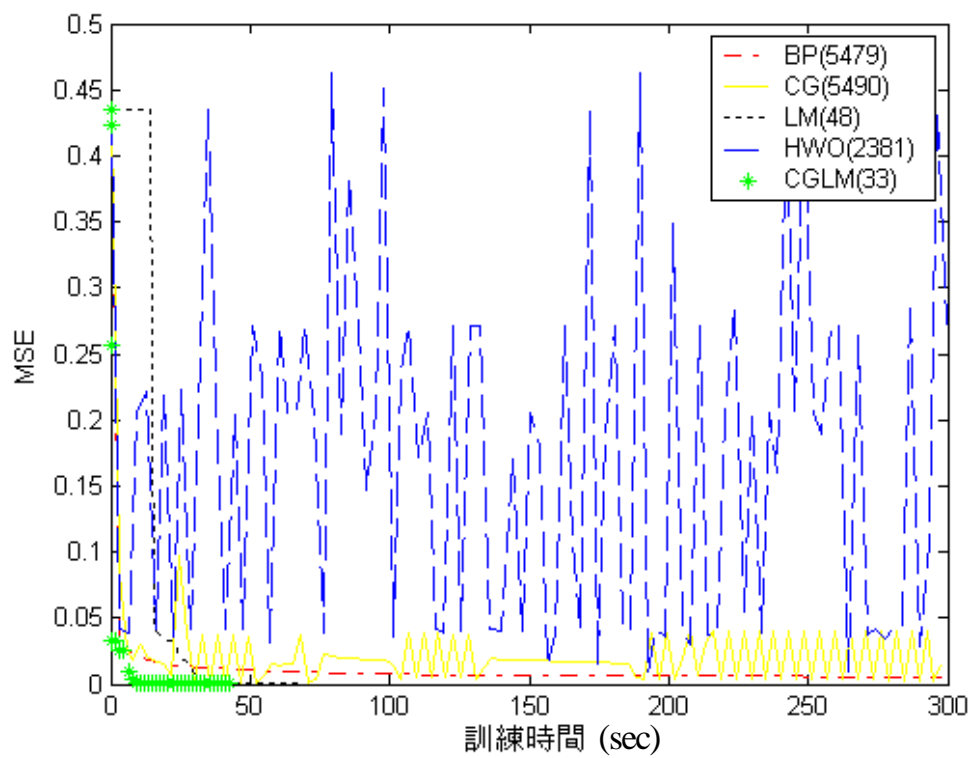


圖 4.13 類型二 EX8 (1-20-1)

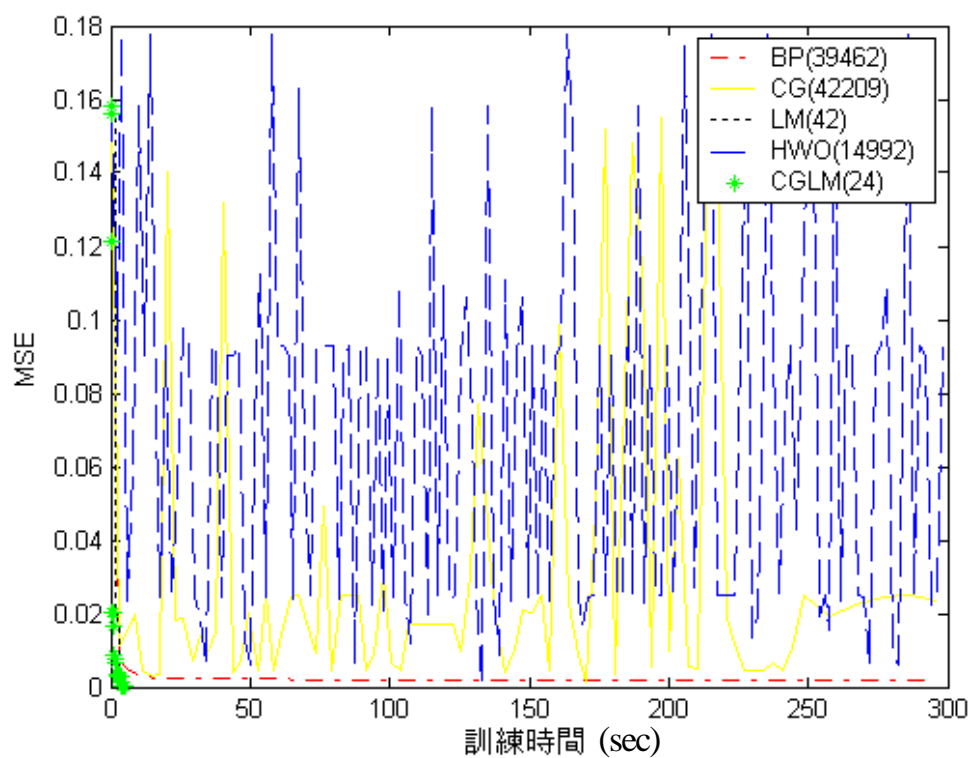


圖 4.14 類型二 EX9 (1-20-1)

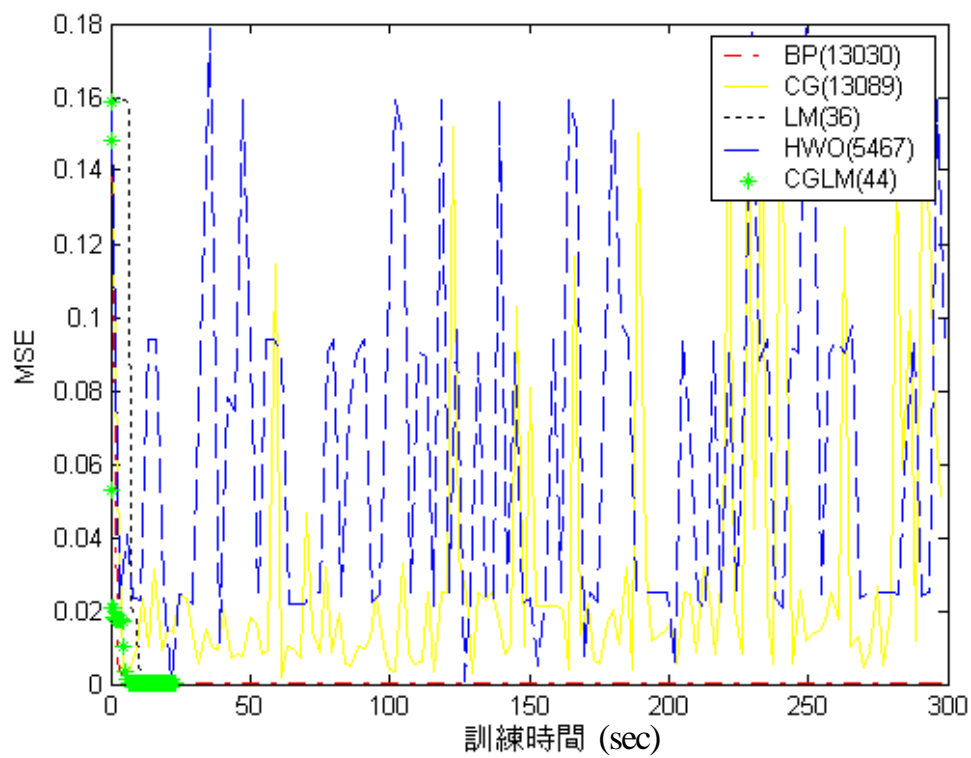


圖 4.15 類型二 EX10 (1-20-1)

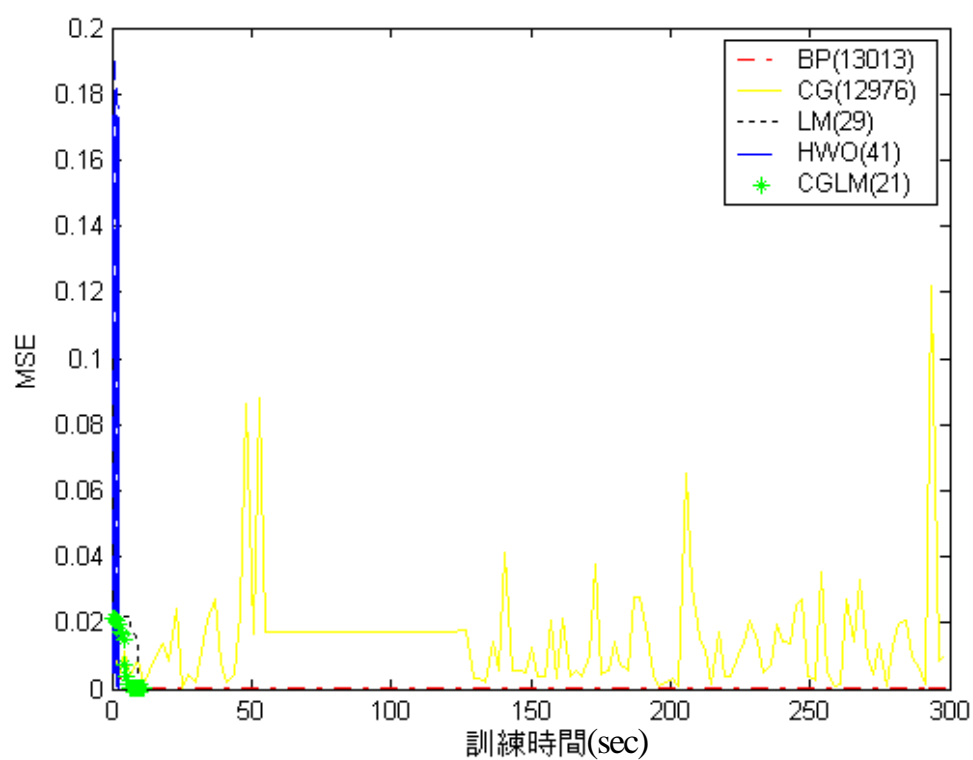


圖 4.16 類型二 EX11(1-20-1)

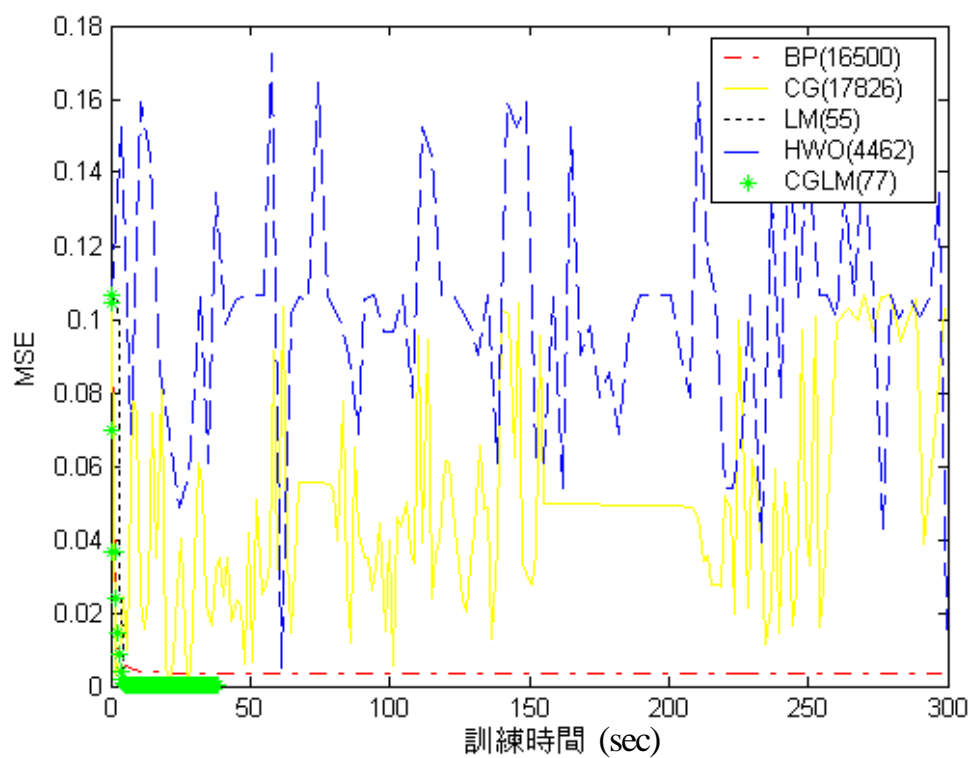


圖 4.17 類型二 EX12 (1-20-1)

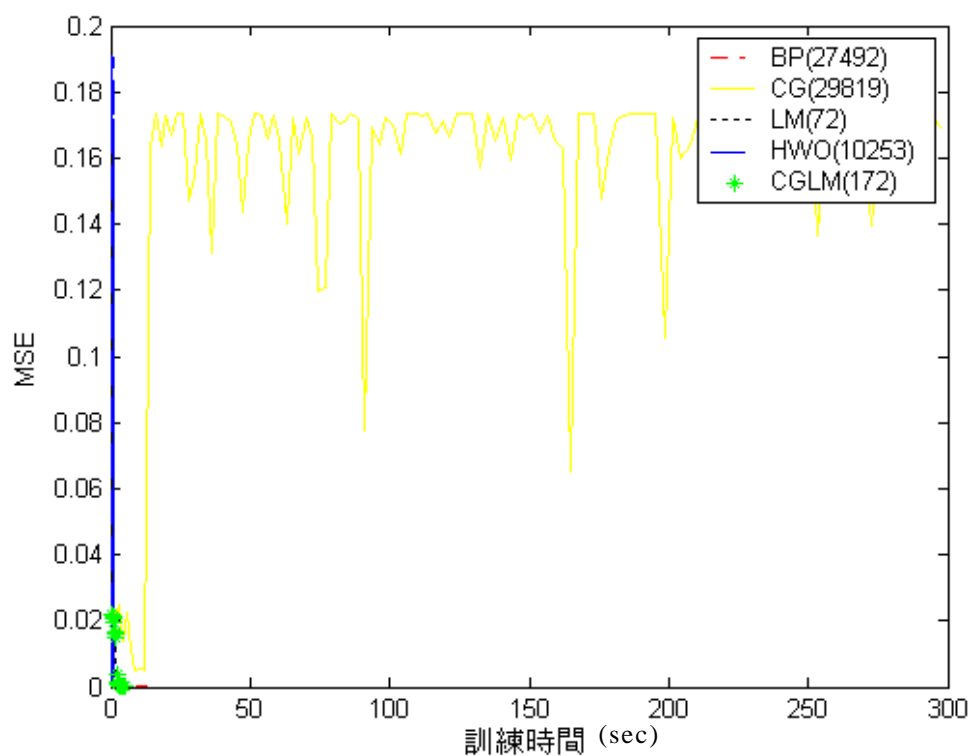


圖 4.18 類型二 EX13 (1-20-1)

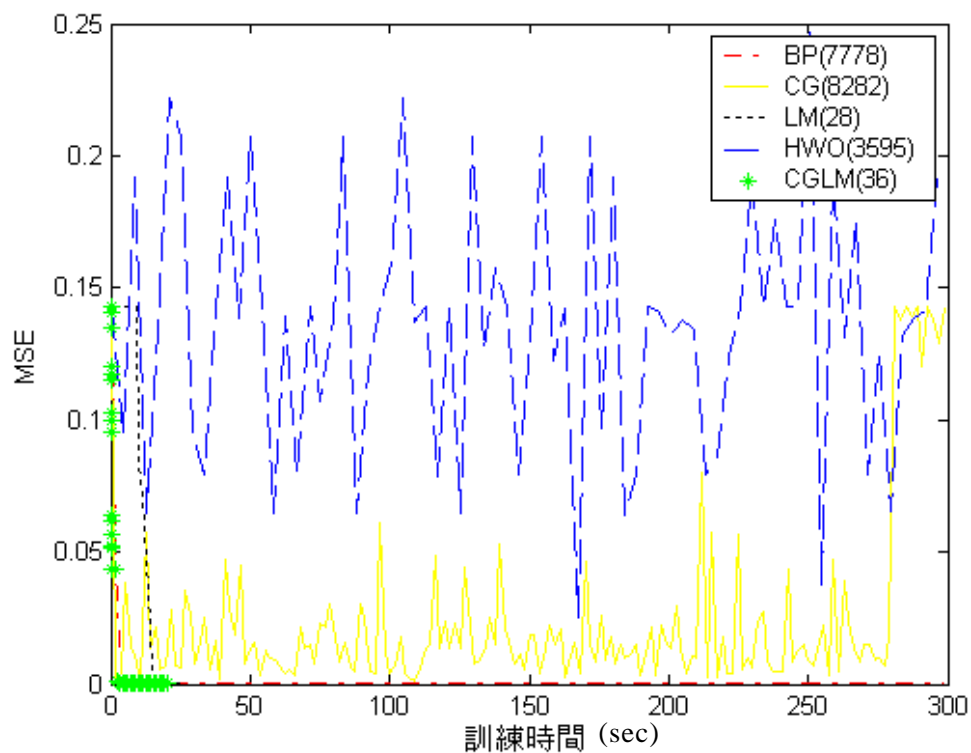


圖 4.19 類型二 EX14 (1-20-1)

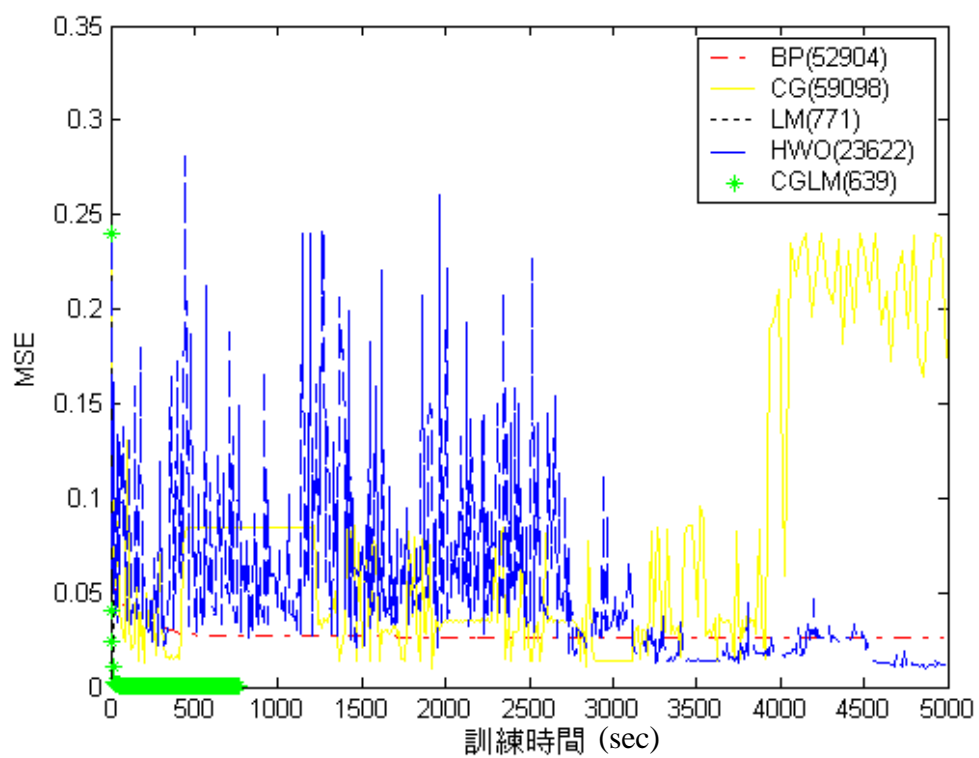


圖 4.20 類型三 (a) EX1 (2-20-1)

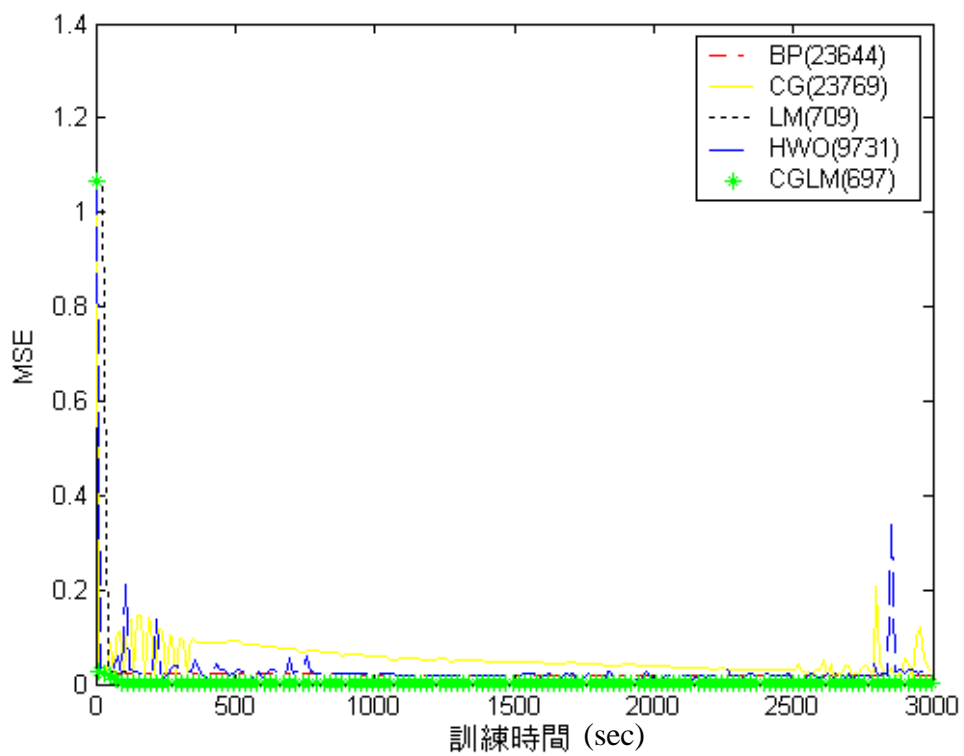


圖 4.21 類型三 (a) EX2 (2-20-1)

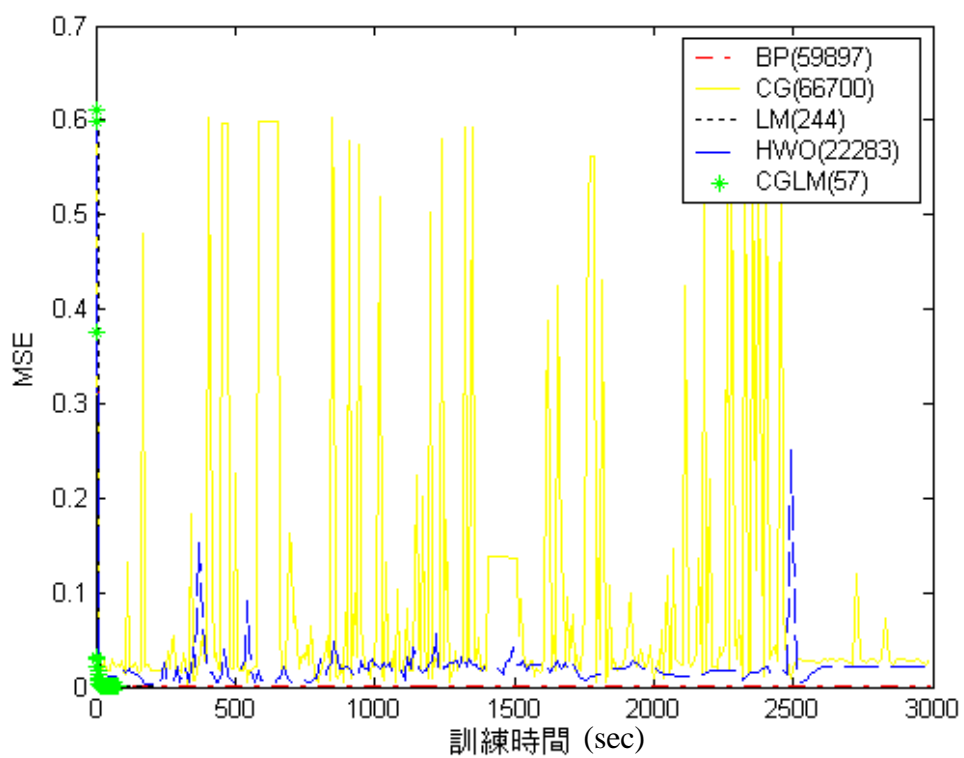


圖 4.22 類型三 (a) EX3 (2-20-1)

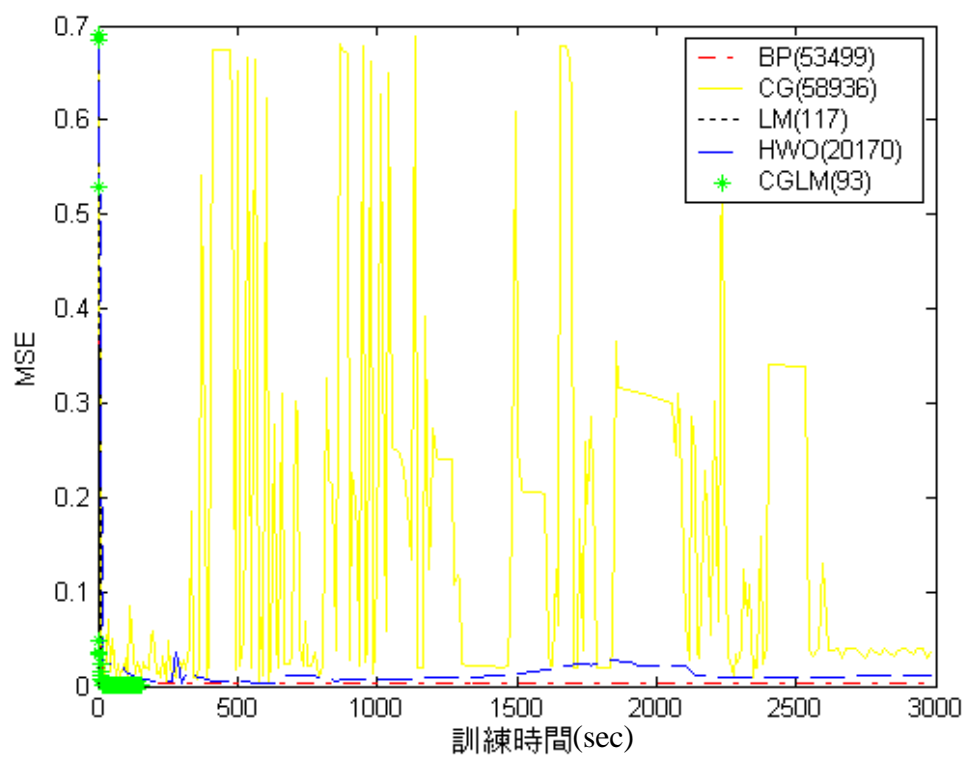


圖 4.23 類型三 (a) EX4 (2-20-1)

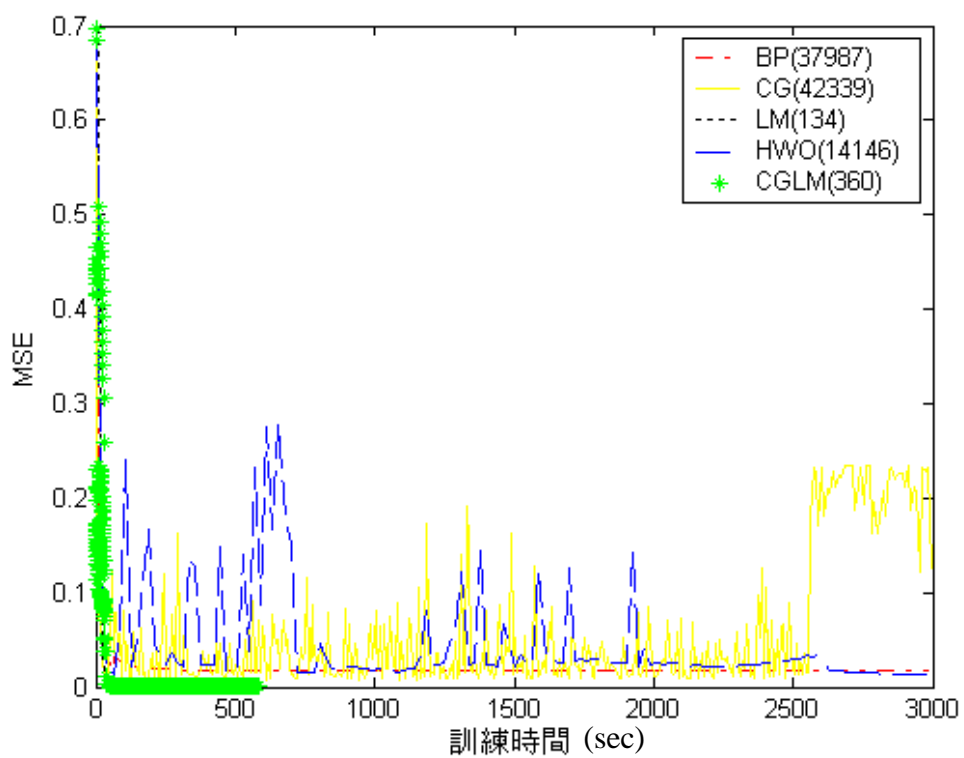


圖 4.24 類型三 (a) EX5 (2-20-1)

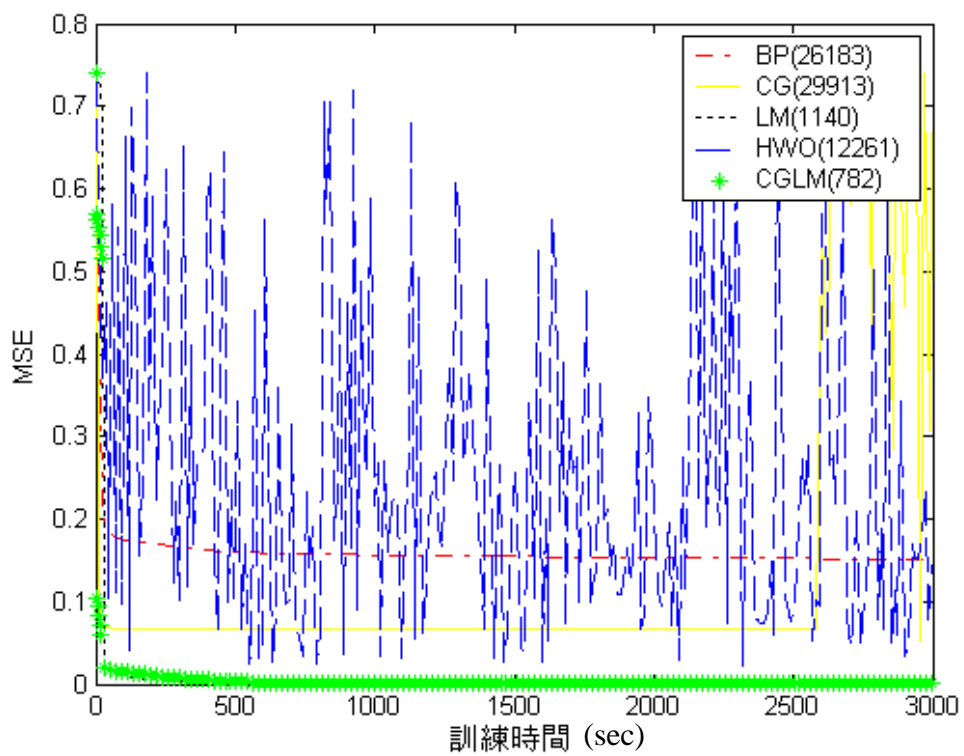


圖 4.25 類型三 (a) EX6 (2-20-1)

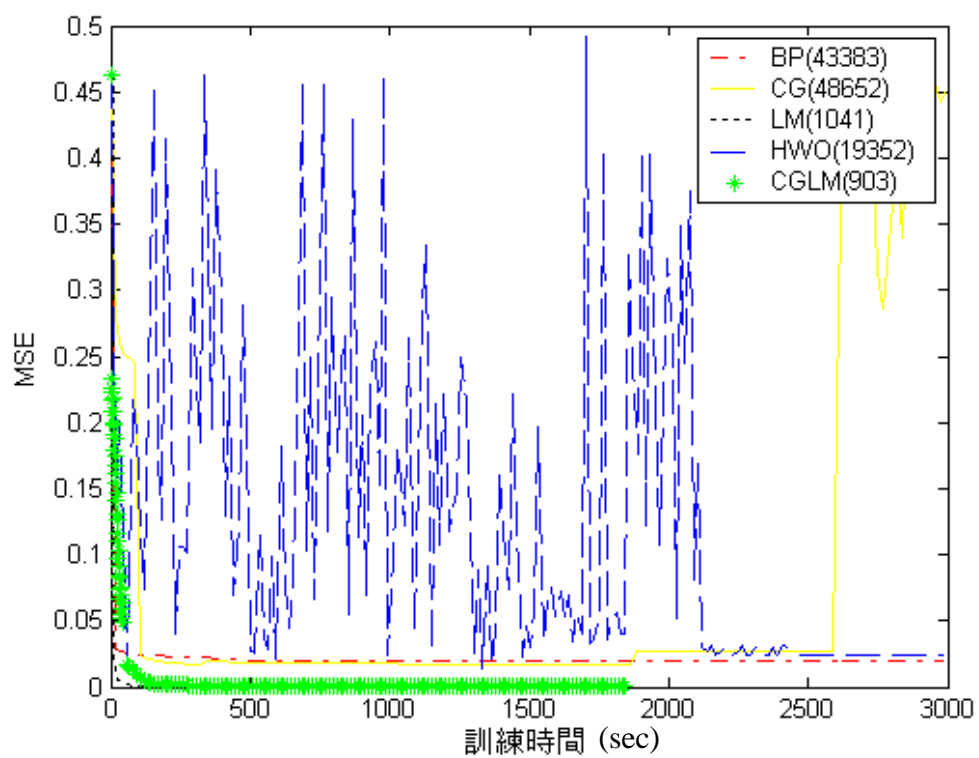


圖 4.26 類型三 (a) EX7 (2-20-1)

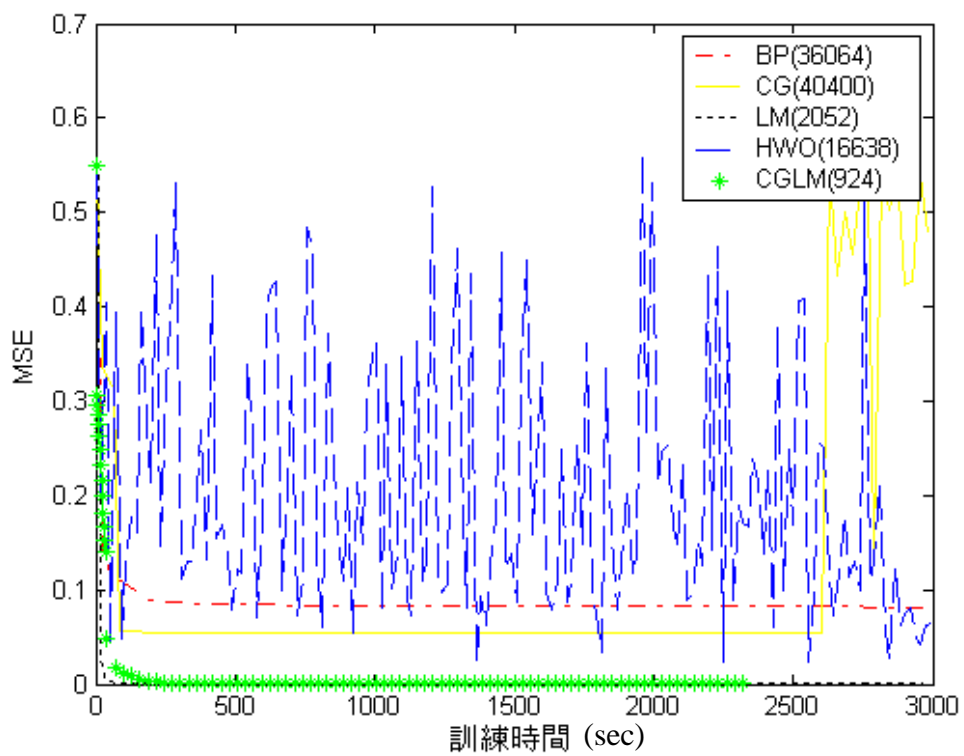


圖 4.27 類型三 (a) EX8 (2-20-1)

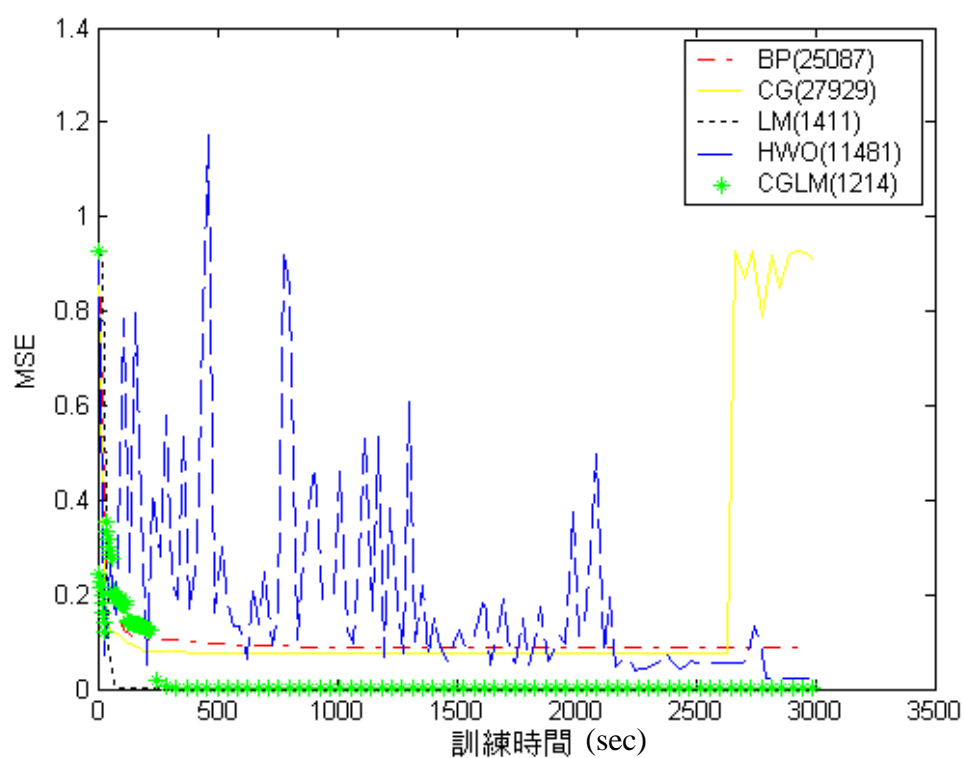


圖 4.28 類型三 (a) EX9 (2-20-1)

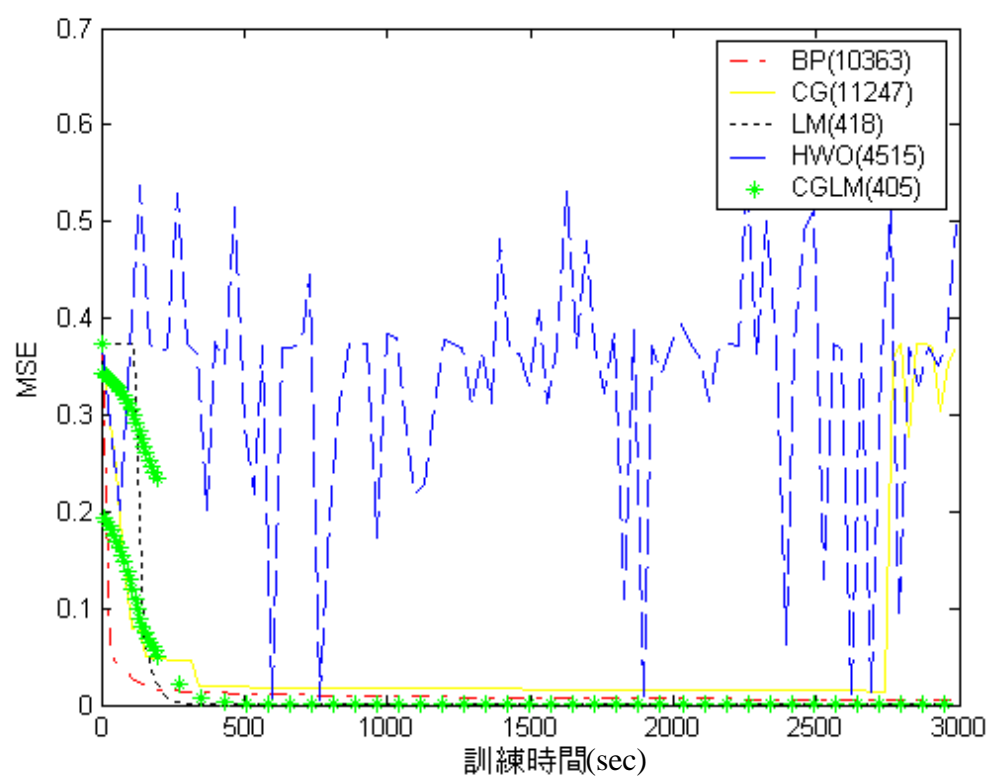


圖 4.29 類型三 (a) EX9 (2-50-1)

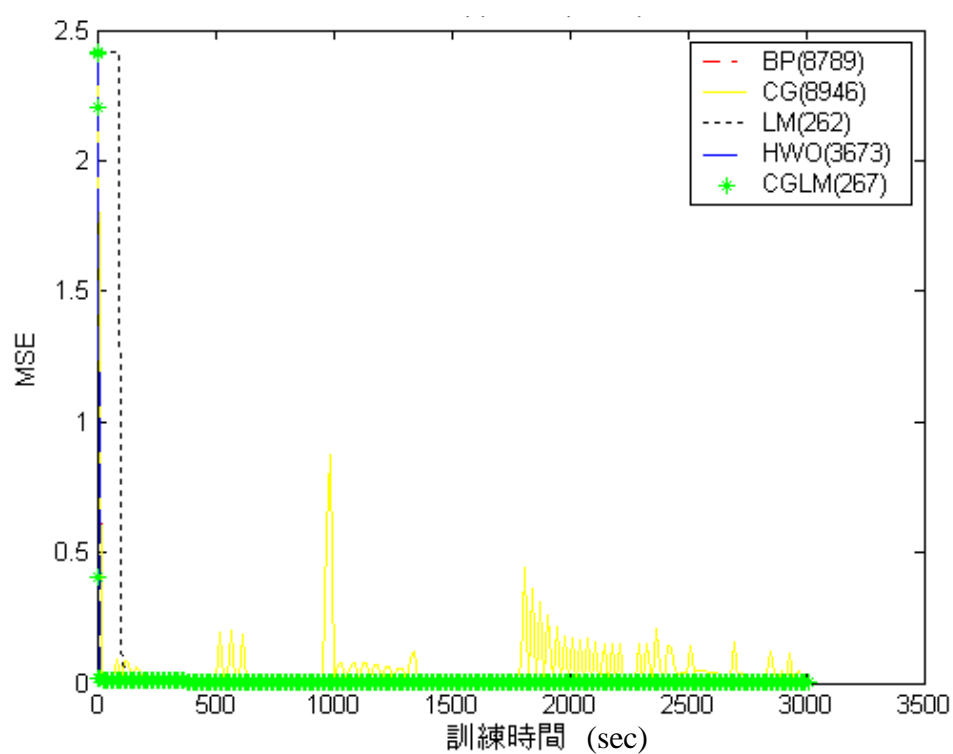


圖 4.30 類型三 (a) EX10 (2-20-1)

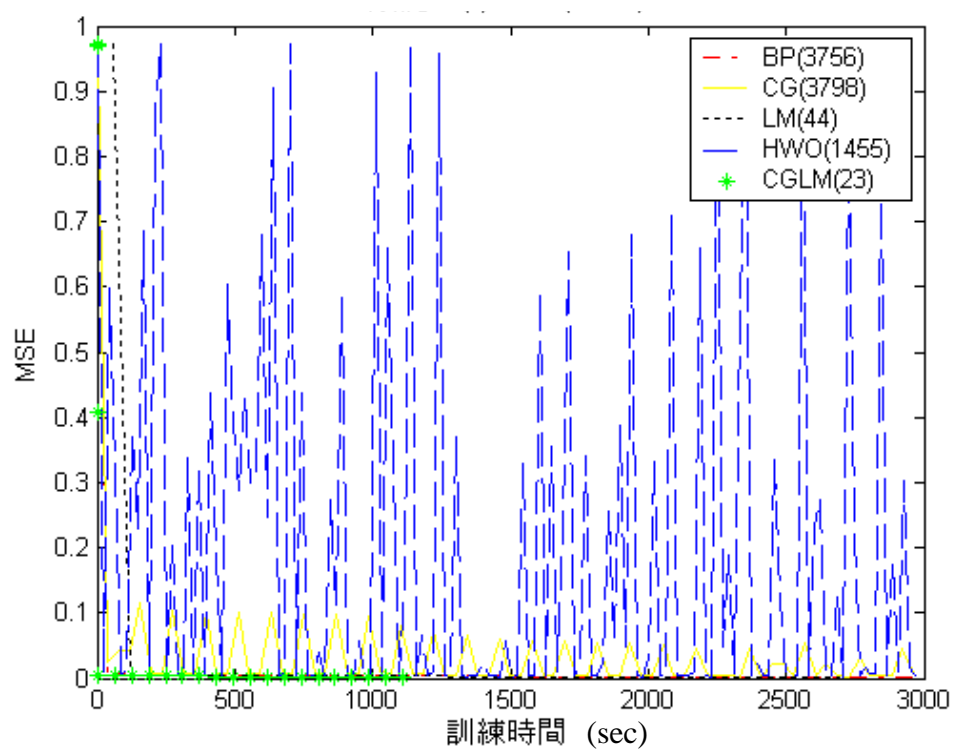


圖 4.31 類型三 (a) EX10 (2-50-1)

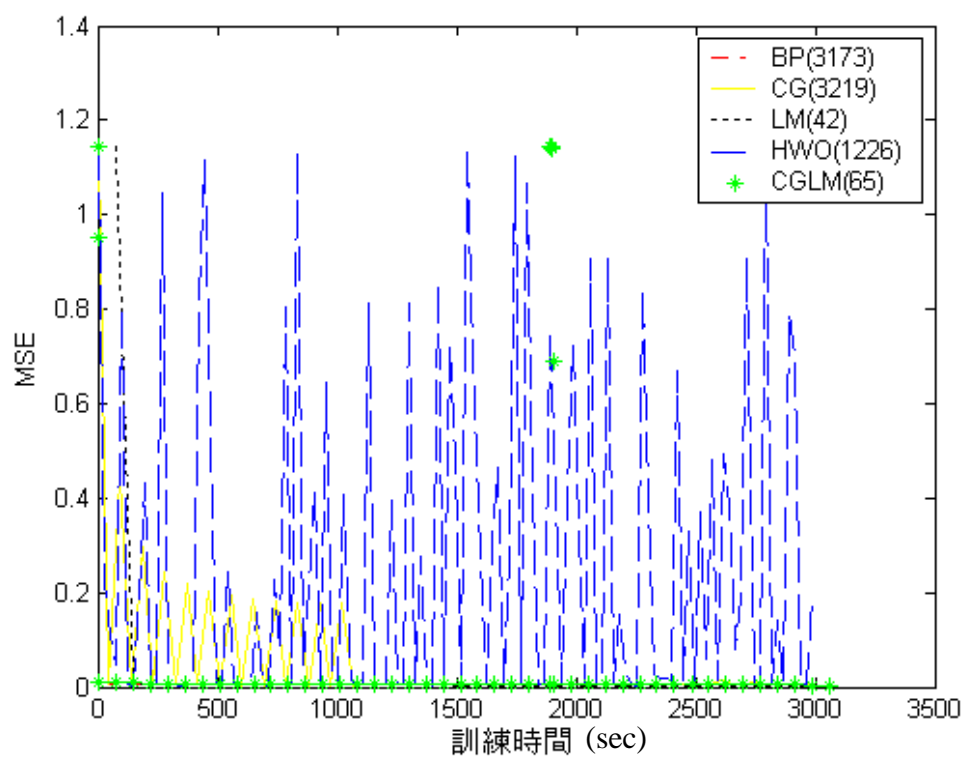


圖 4.32 類型三 (a) EX11 (2-50-1)

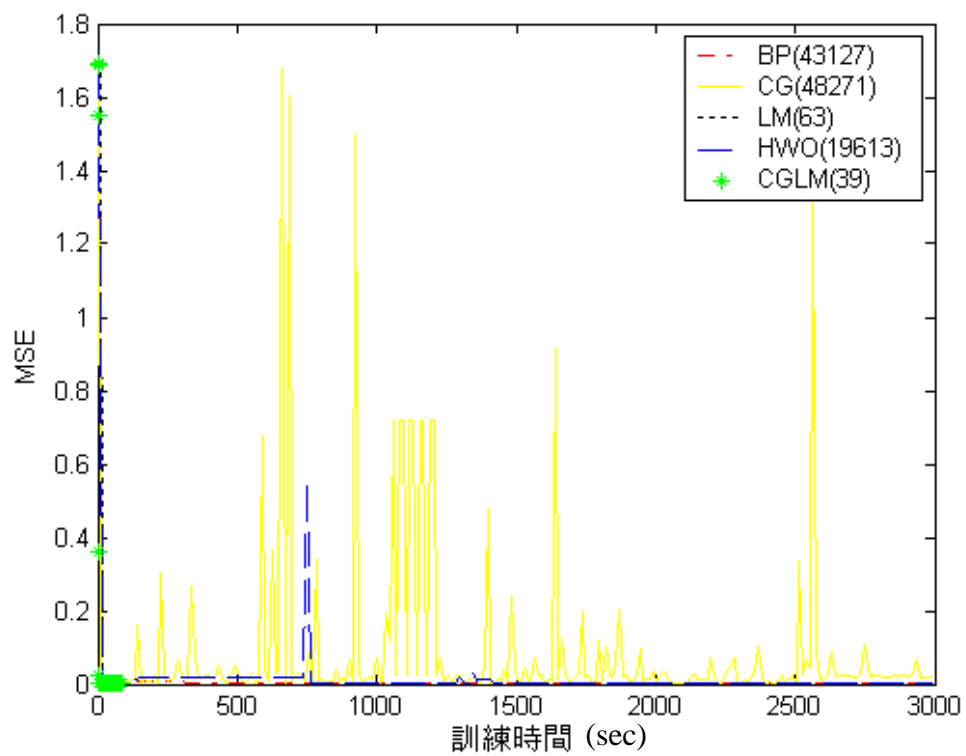


圖 4.33 類型三 (a) EX12 (2-20-1)

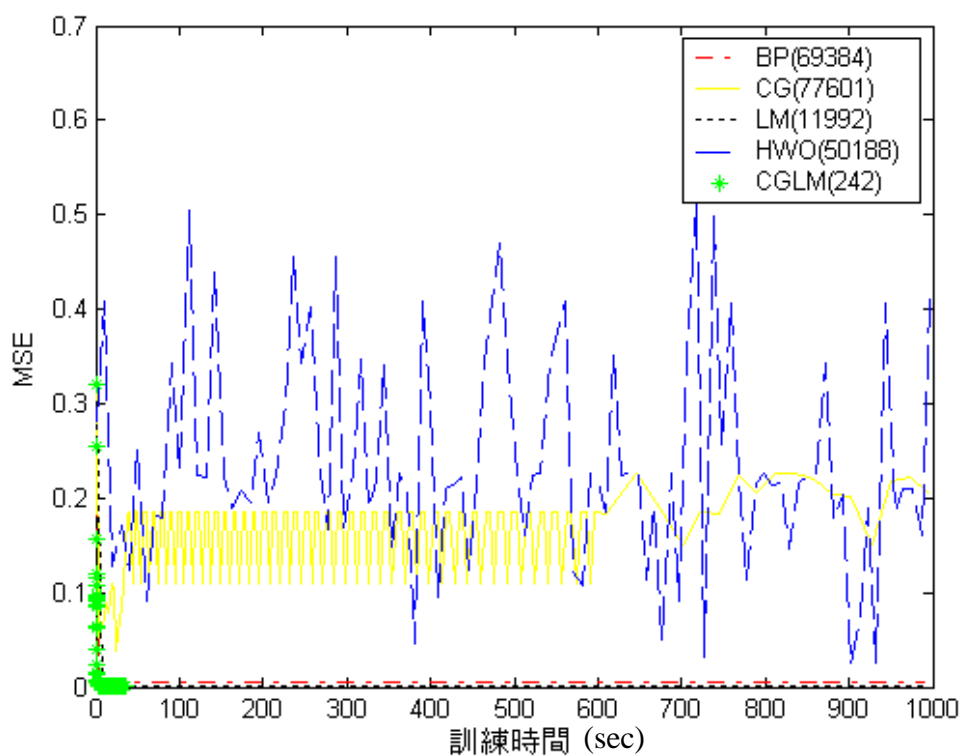


圖 4.34 類型三 (b) EX1 (1-8-2)

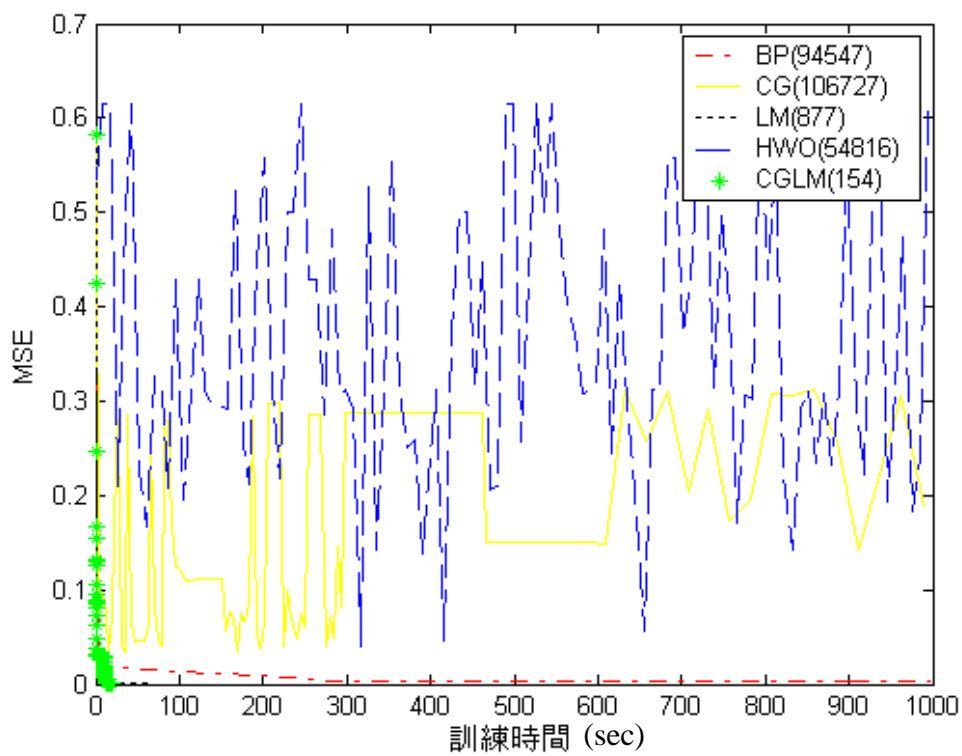


圖 4.35 類型三 (b) EX2 (1-8-2)

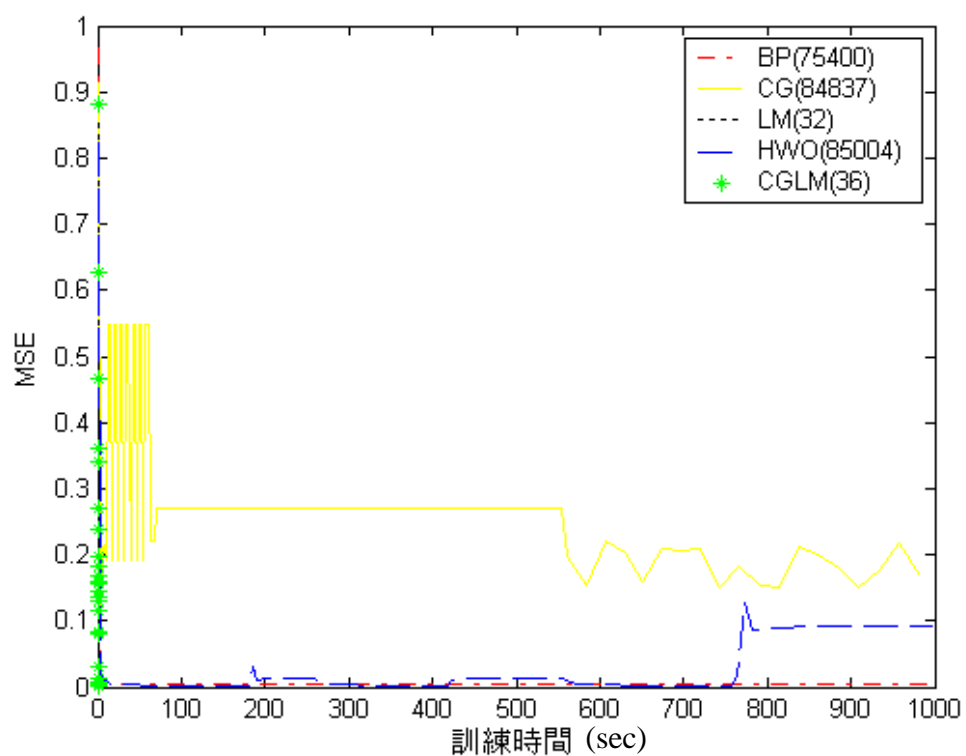


圖 4.36 類型三 (b) EX3 (1-3-2)

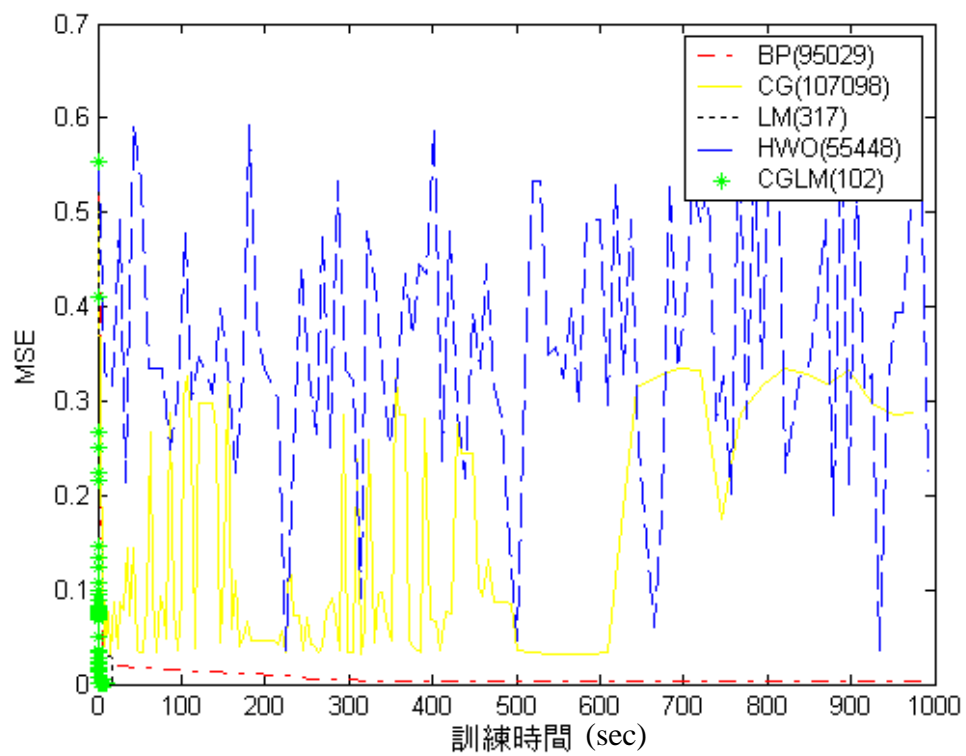


圖 4.37 類型三 (b) EX4 (1-8-2)

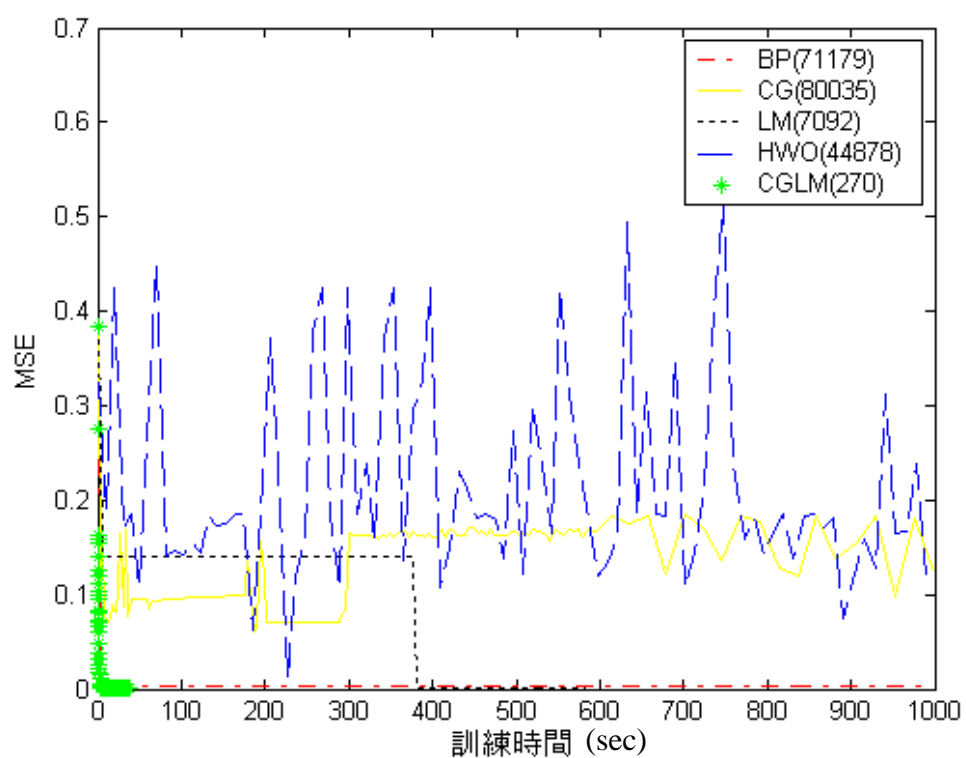


圖 4.38 類型三 (b) EX5 (1-8-2)

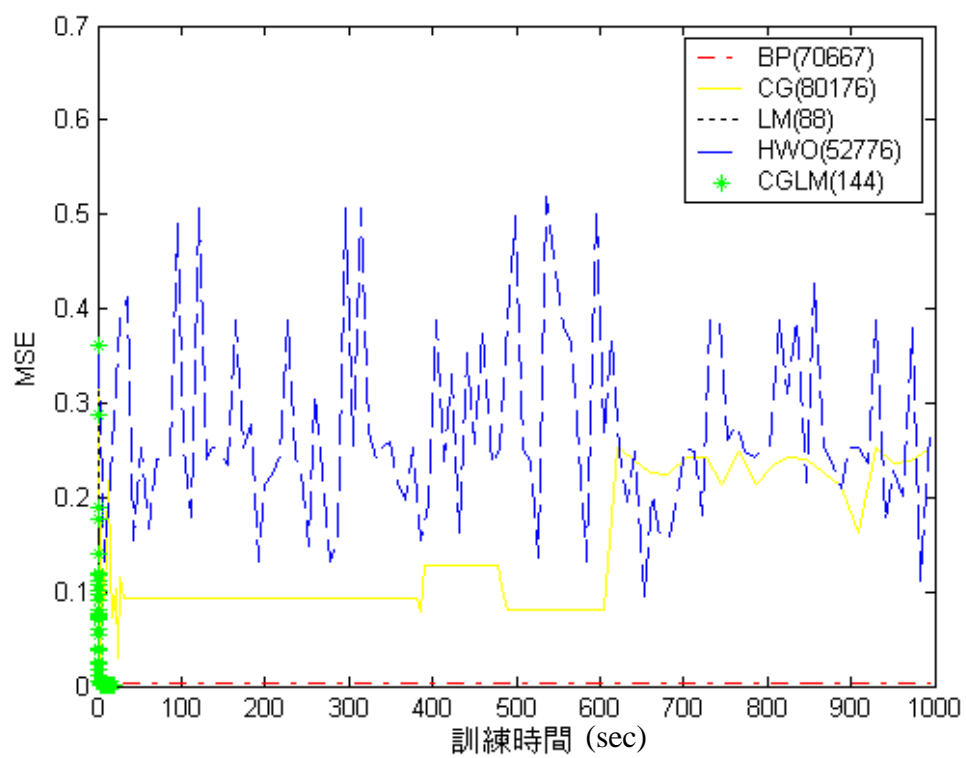


圖 4.39 類型三 (b) EX6 (1-8-2)

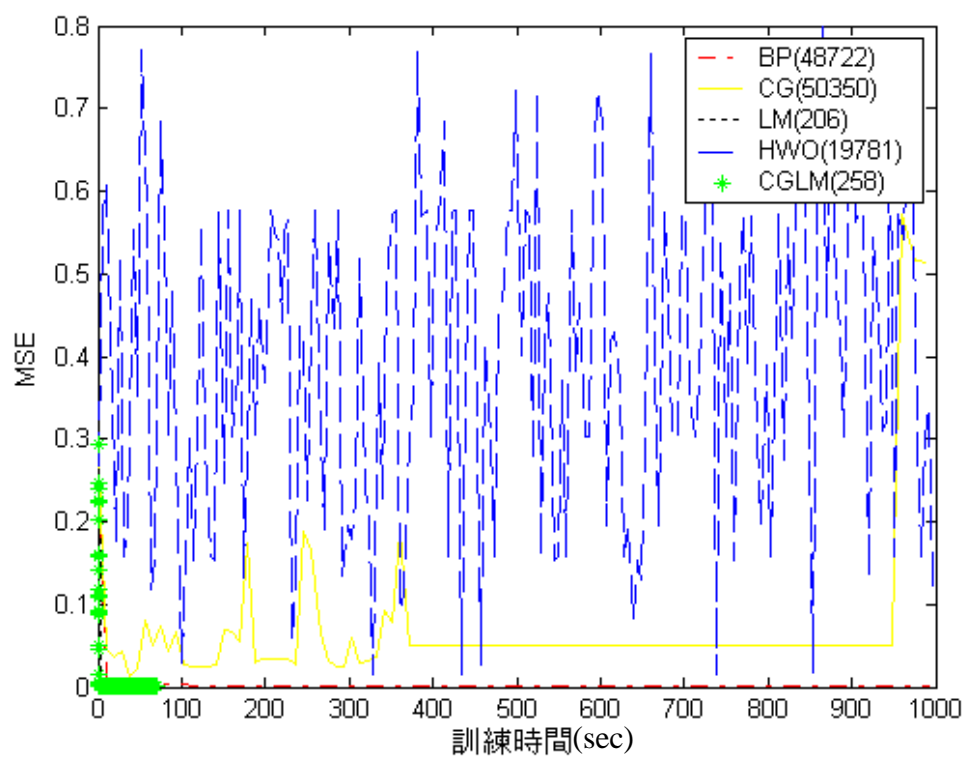


圖 4.40 類型三 (b) EX7 (1-8-2)

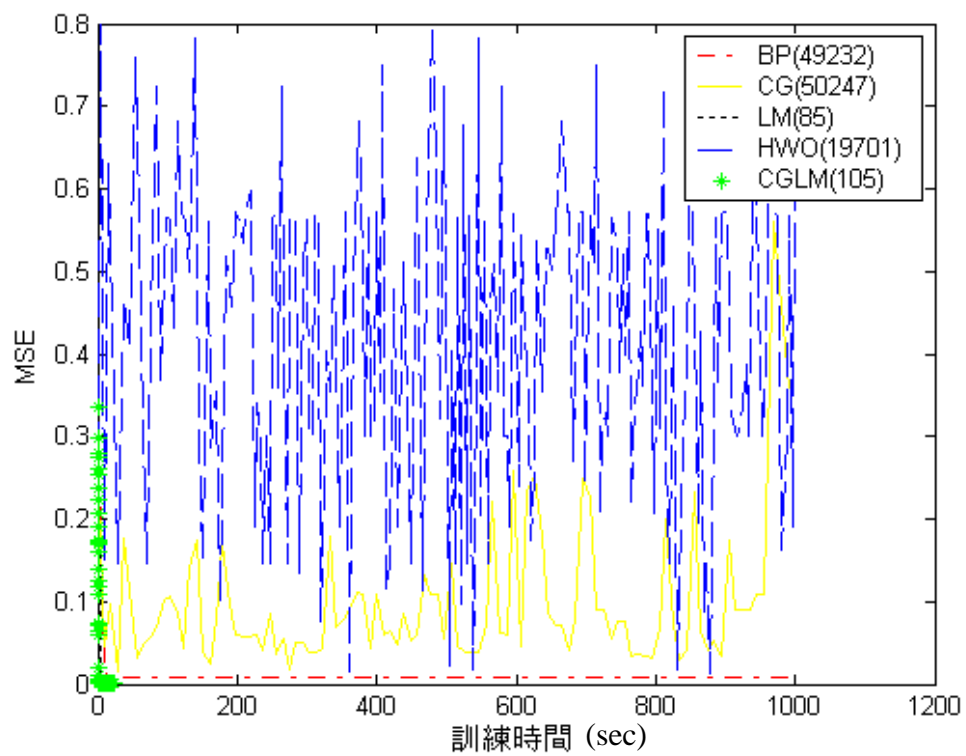


圖 4.41 類型三 (b) EX8 (1-8-2)

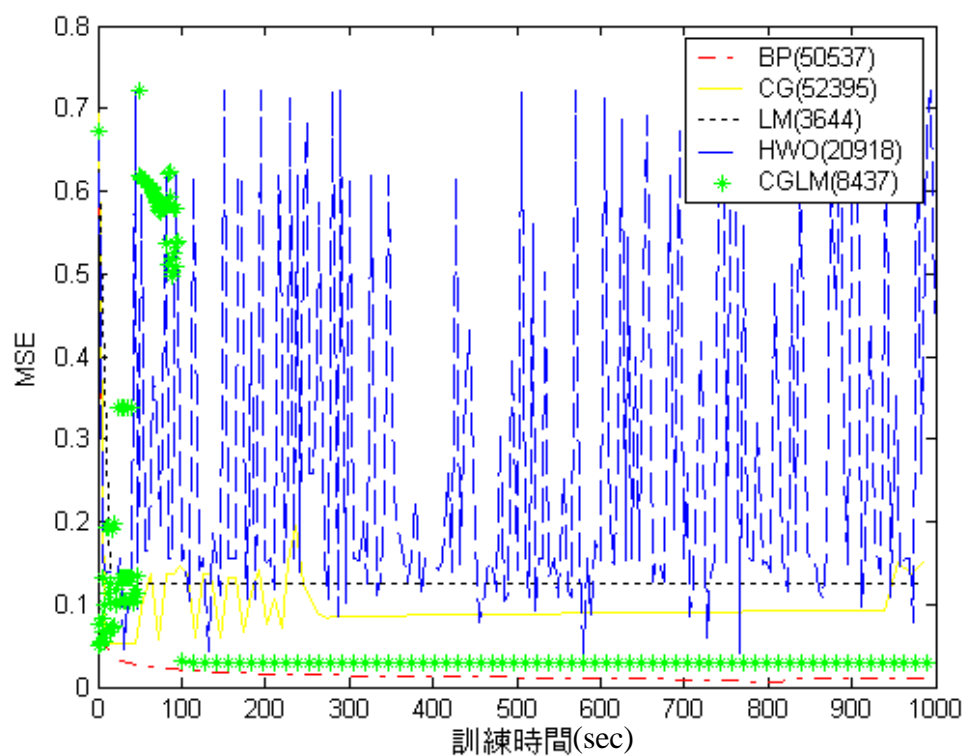


圖 4.42 類型三 (b) EX9 (1-8-2)

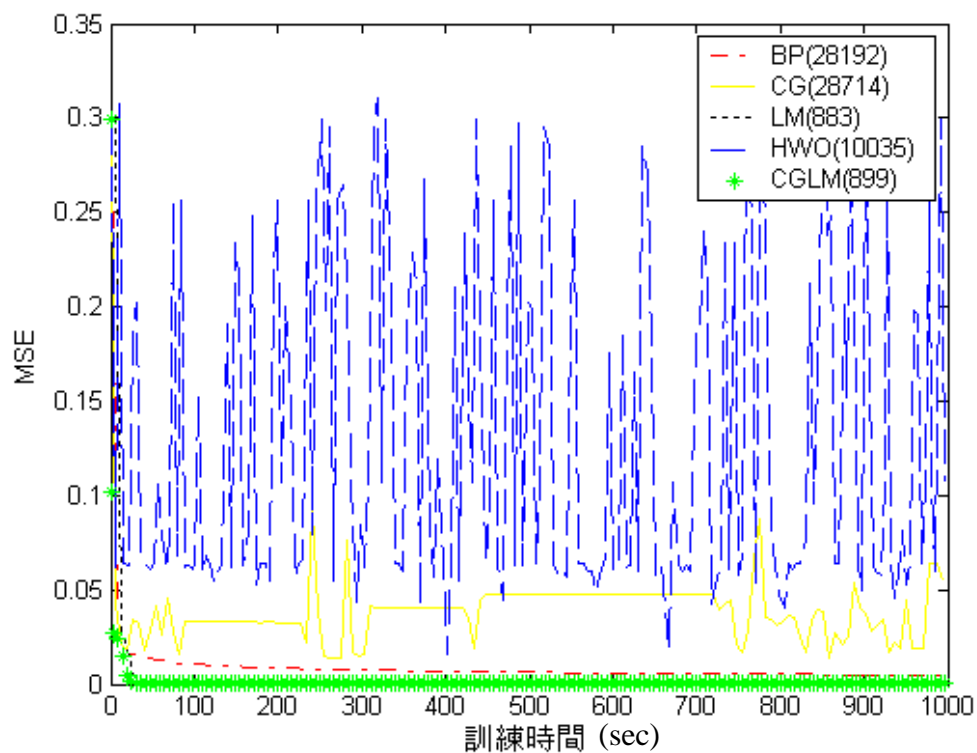


圖 4.43 類型三 (b) EX9 (1-20-2)

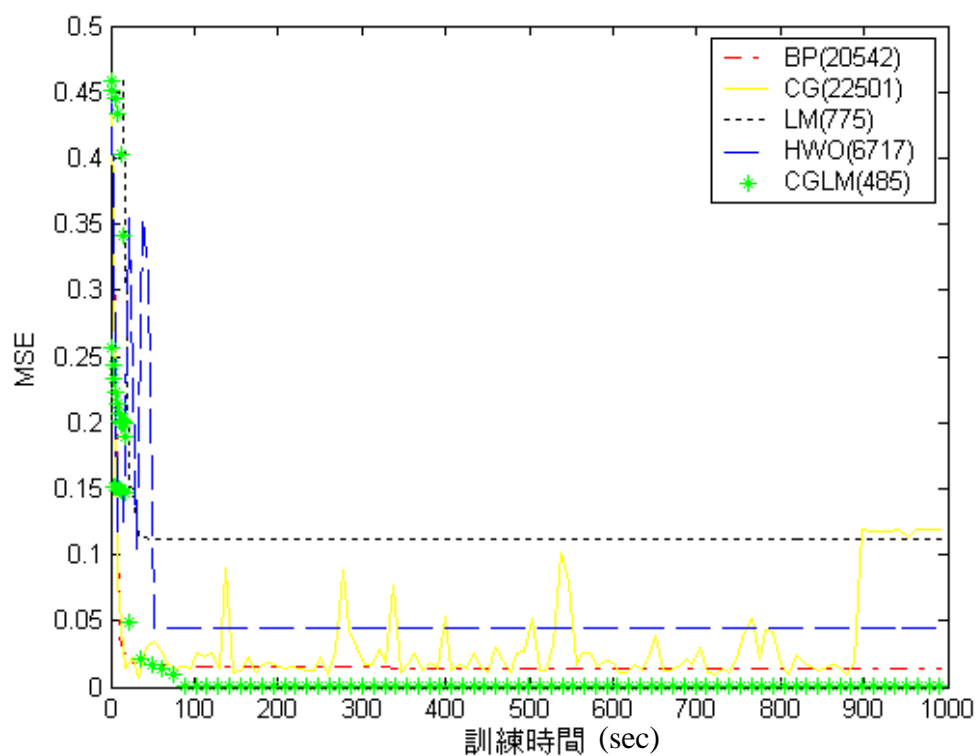


圖 4.44 類型三 (b) EX10 (1-40-2)

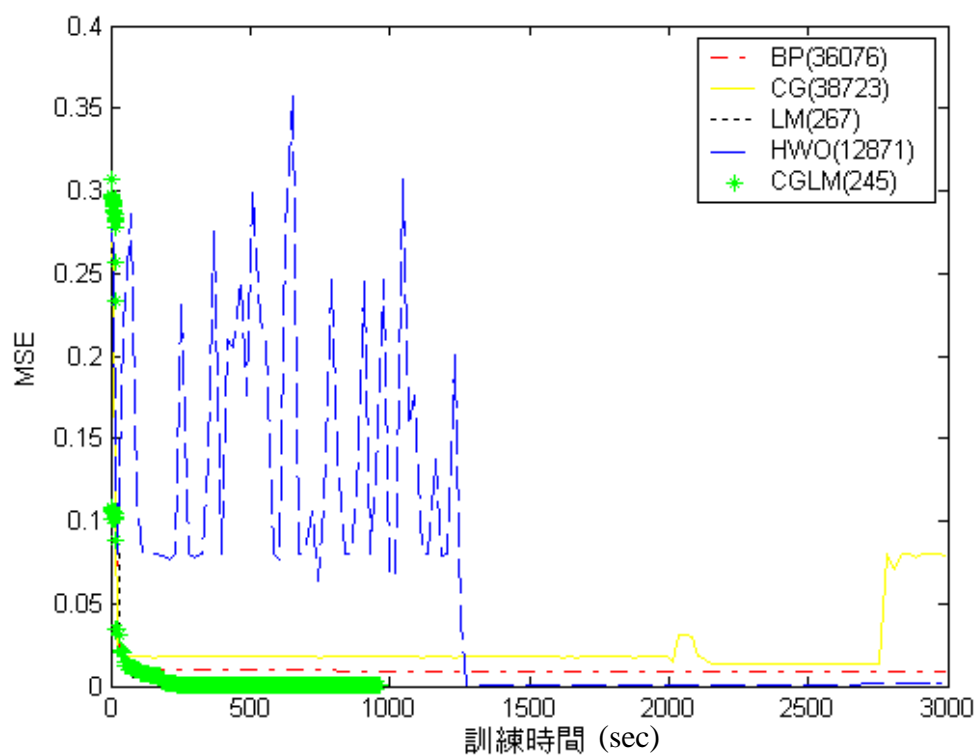


圖 4.45 類型三 (b) EX10 (1-60-2)

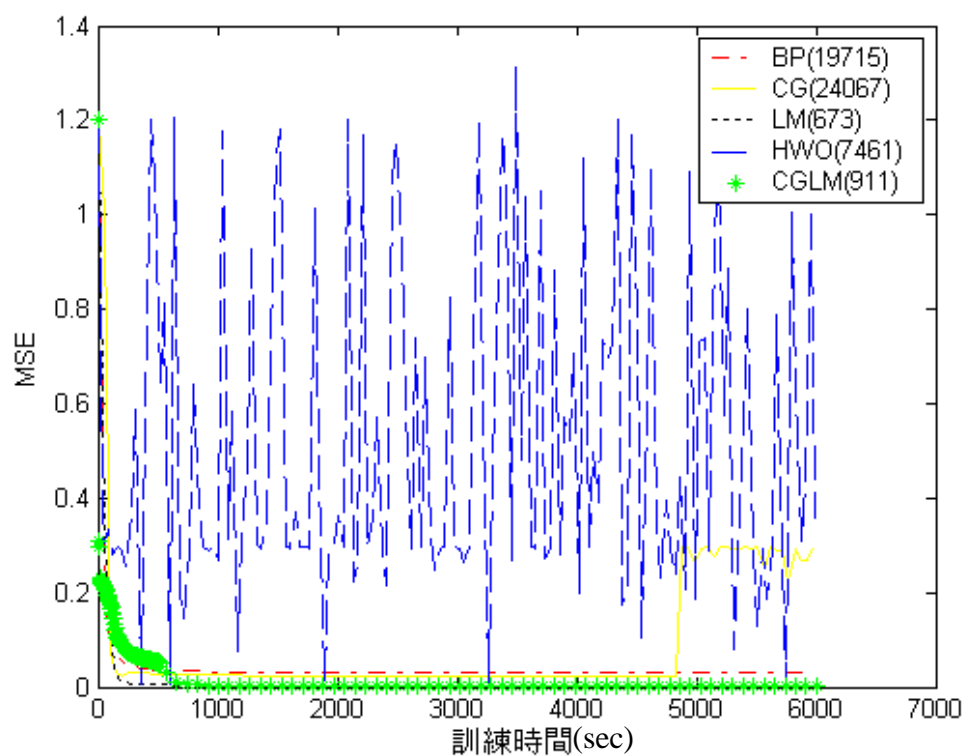


圖 4.46 類型三 (c) EX1 (2-50-2)

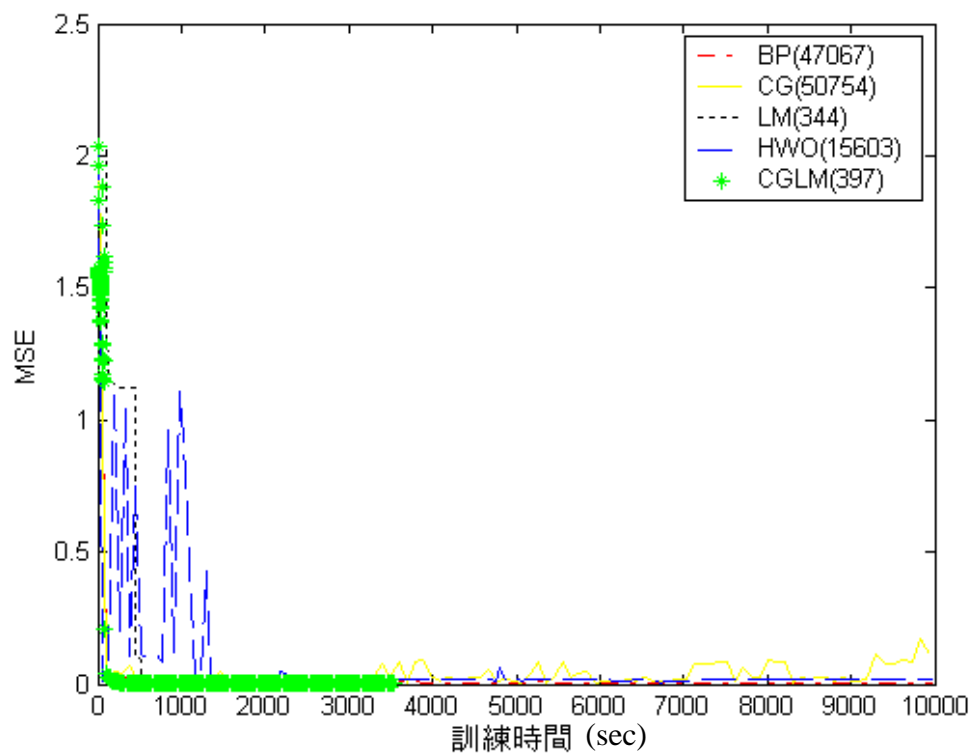


圖 4.47 類型三 (c) EX2 (2-40-2)

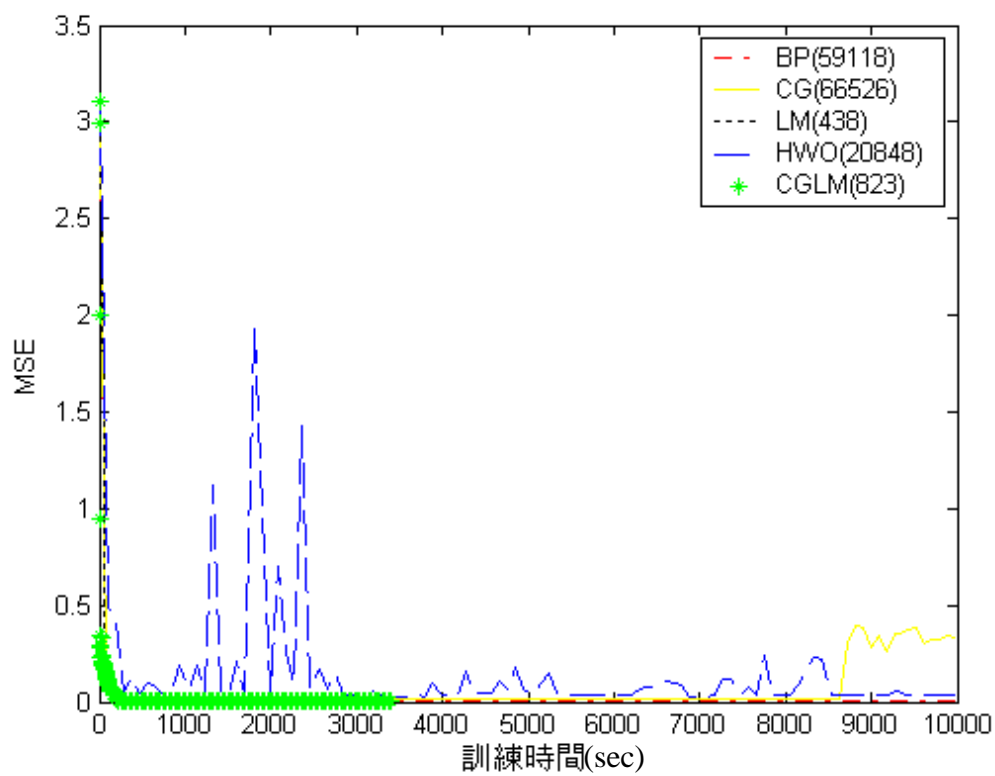


圖 4.48 類型三 (c) EX3 (2-20-2)

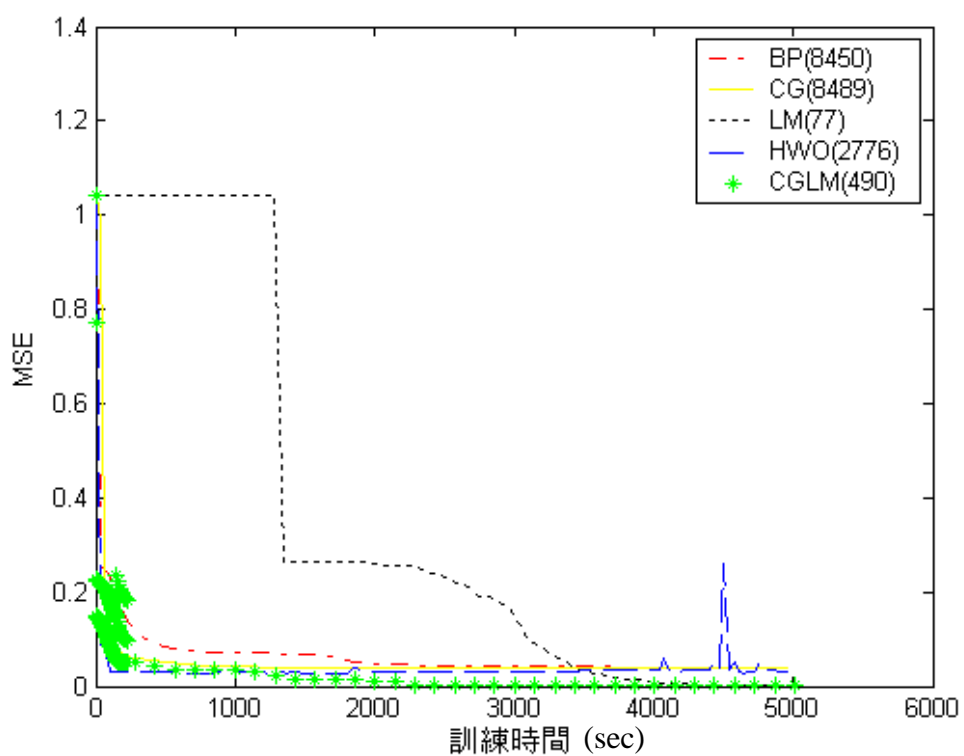


圖 4.49 類型三 (c) EX4 (2-40-2)

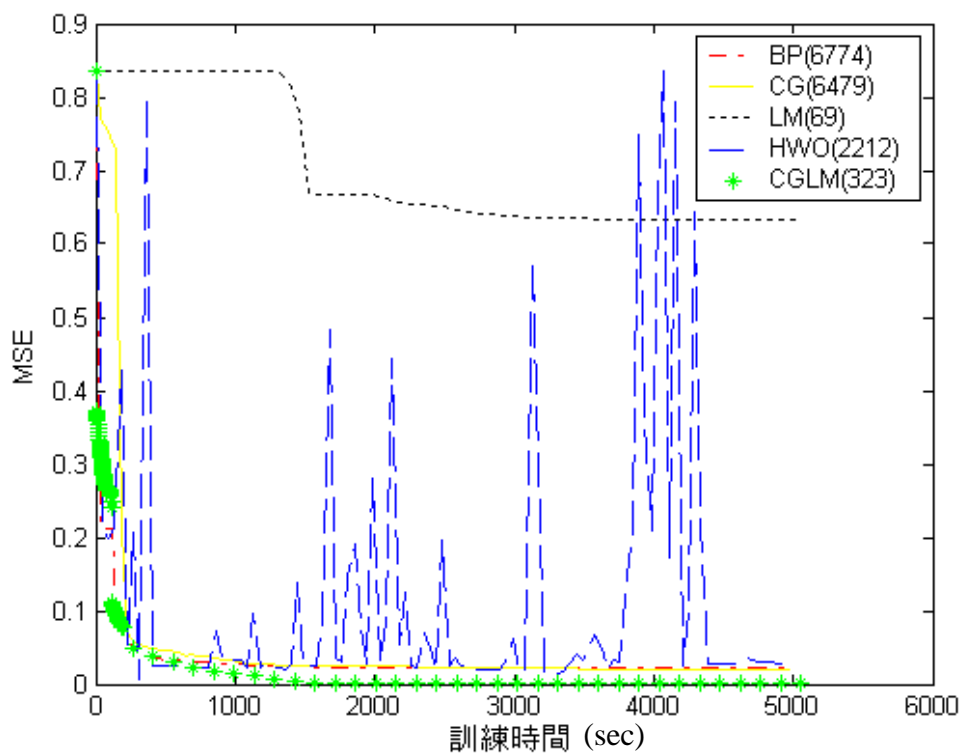


圖 4.50 類型三 (c) EX4 (2-50-2)

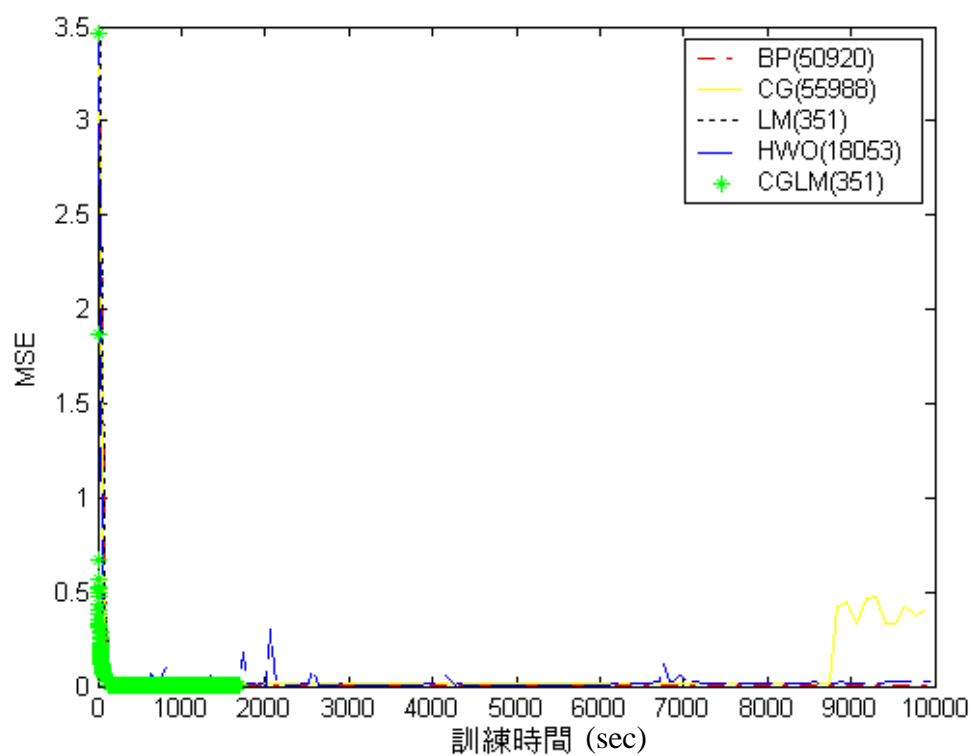


圖 4.51 類型三 (c) EX5 (2-20-2)

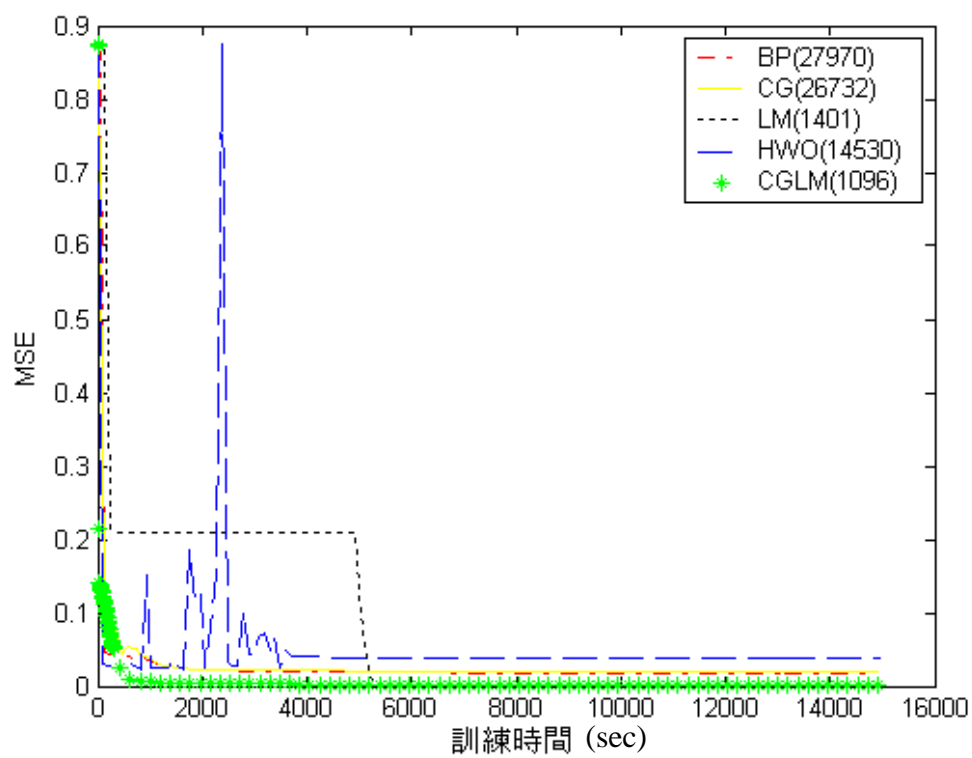


圖 4.52 類型三 (c) EX6 (2-50-2)

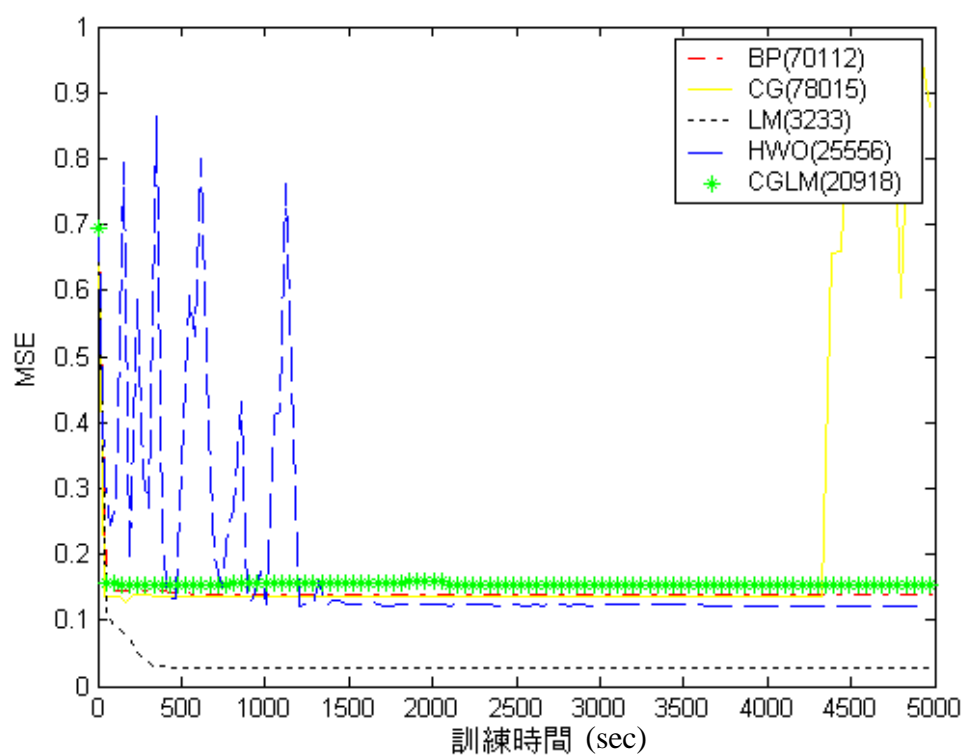


圖 4.53 類型三 (c) EX7 (2-20-2)

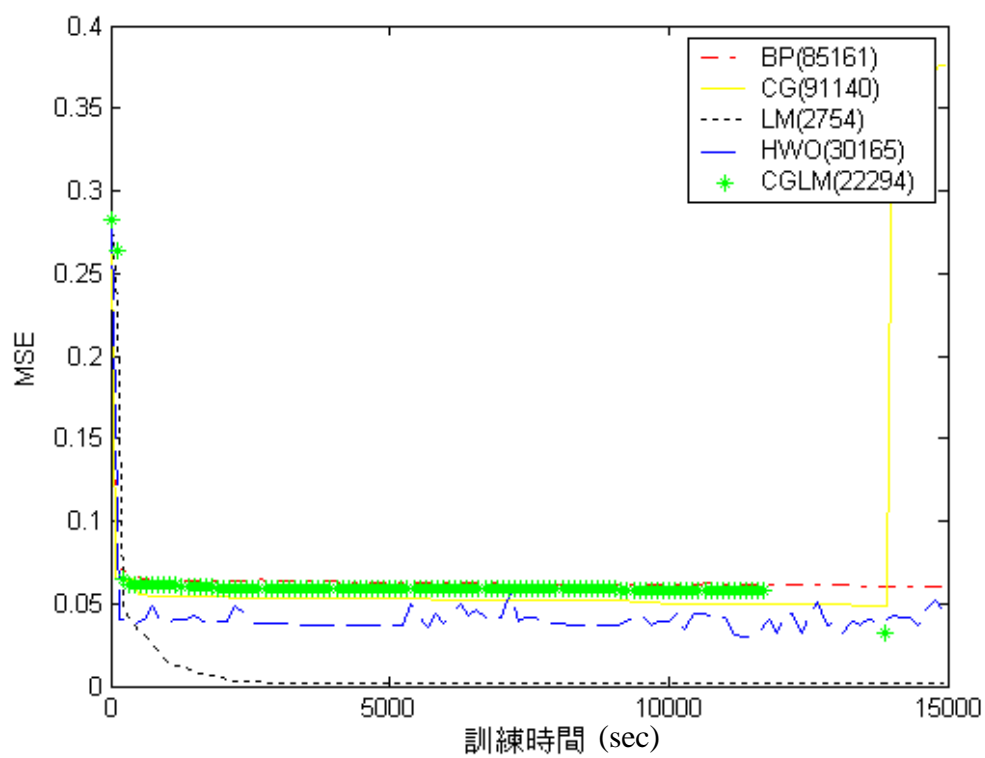


圖 4.54 類型三 (c) EX7 (2-50-2)

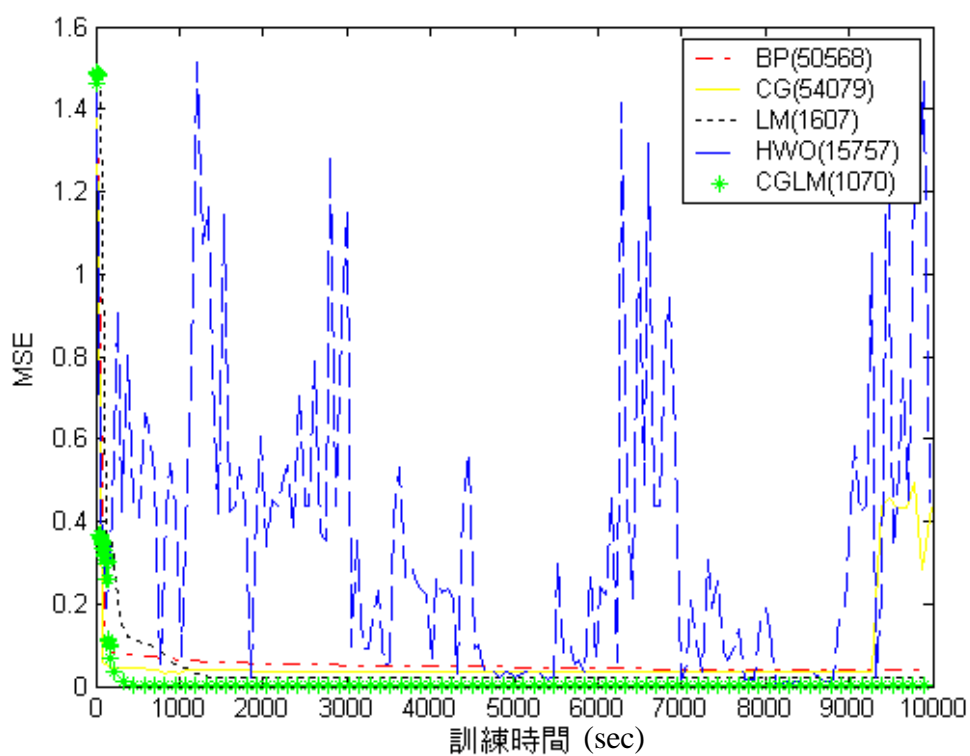


圖 4.55 類型三 (c) EX8 (2-40-2)

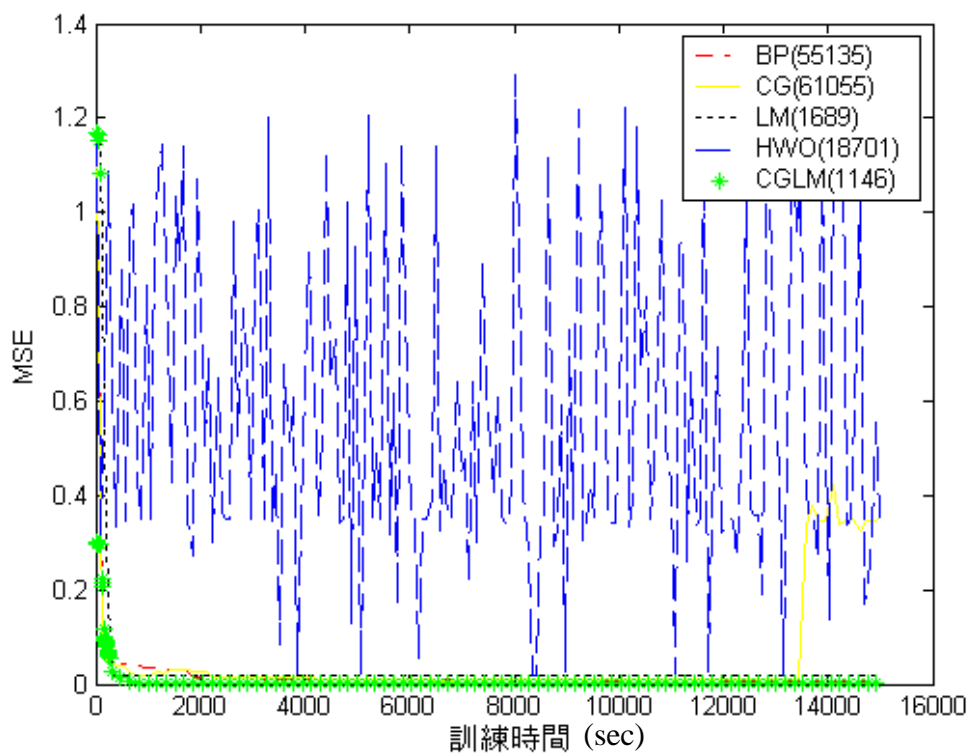


圖 4.56 類型三 (c) EX8 (2-50-2)

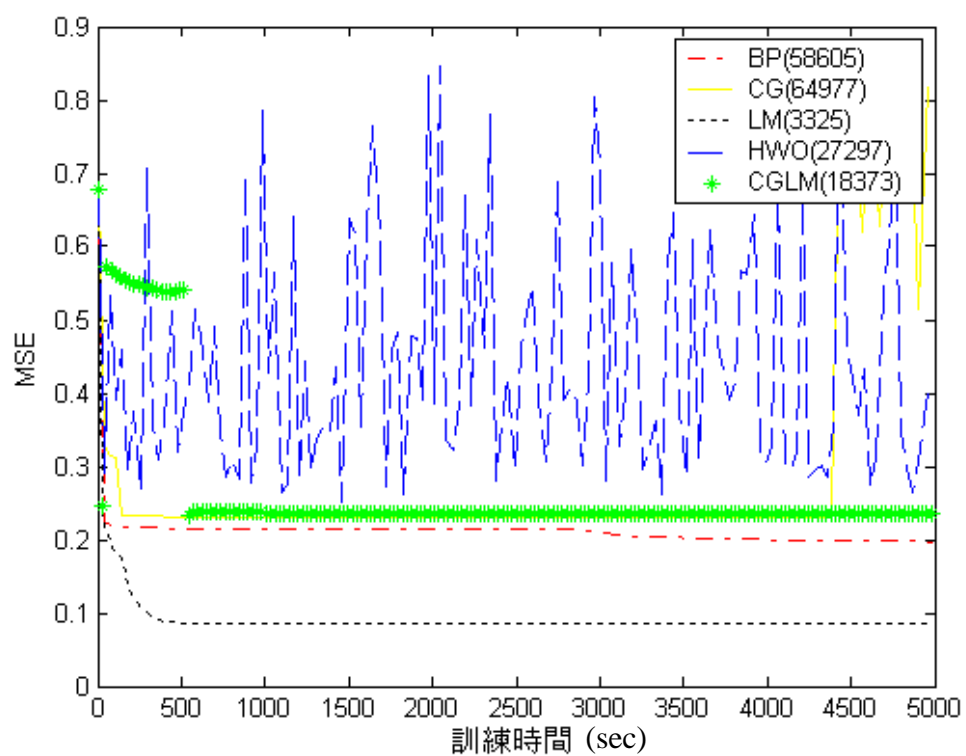


圖 4.57 類型三 (d) EX1 (2-20-1)

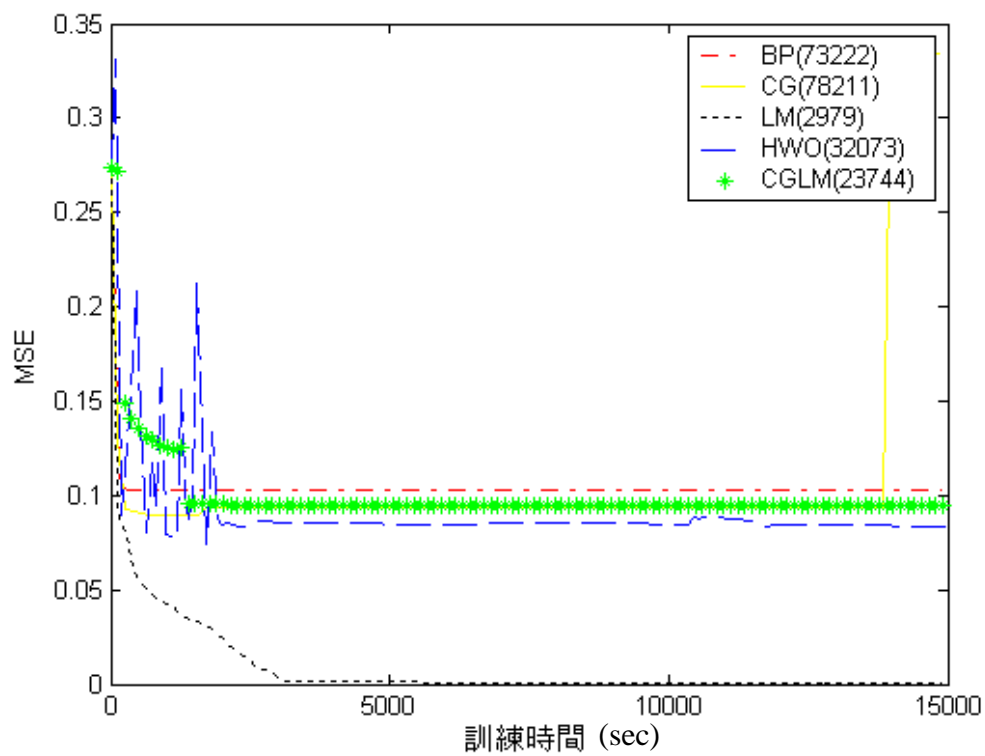


圖 4.58 類型三 (d) EX1 (2-50-1)

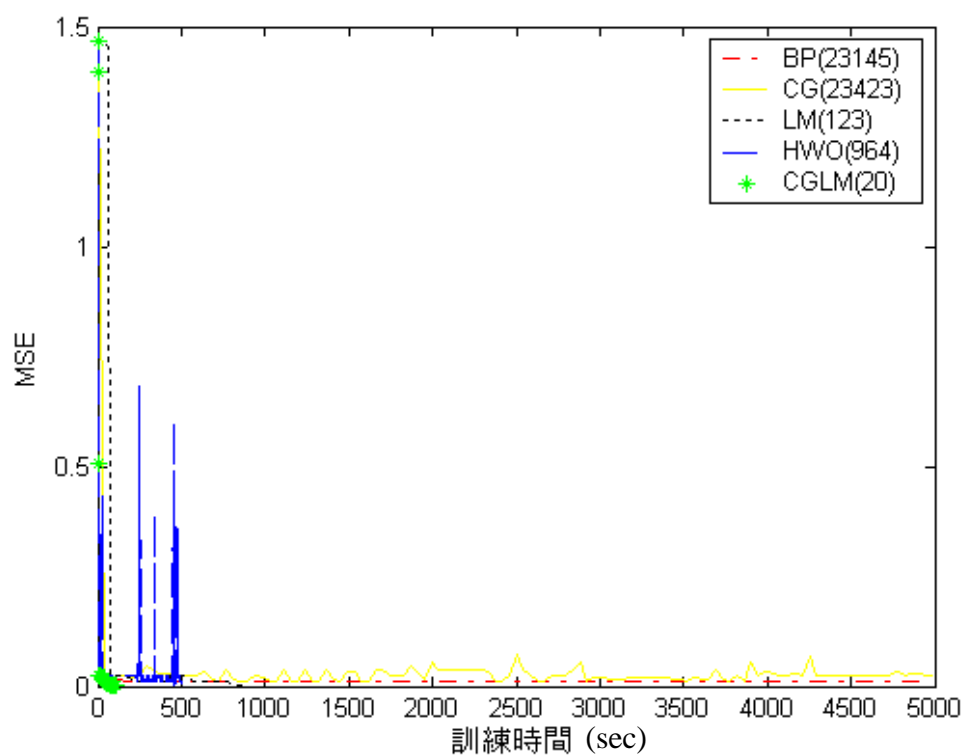


圖 4.59 類型三 (d) EX2 (2-20-1)

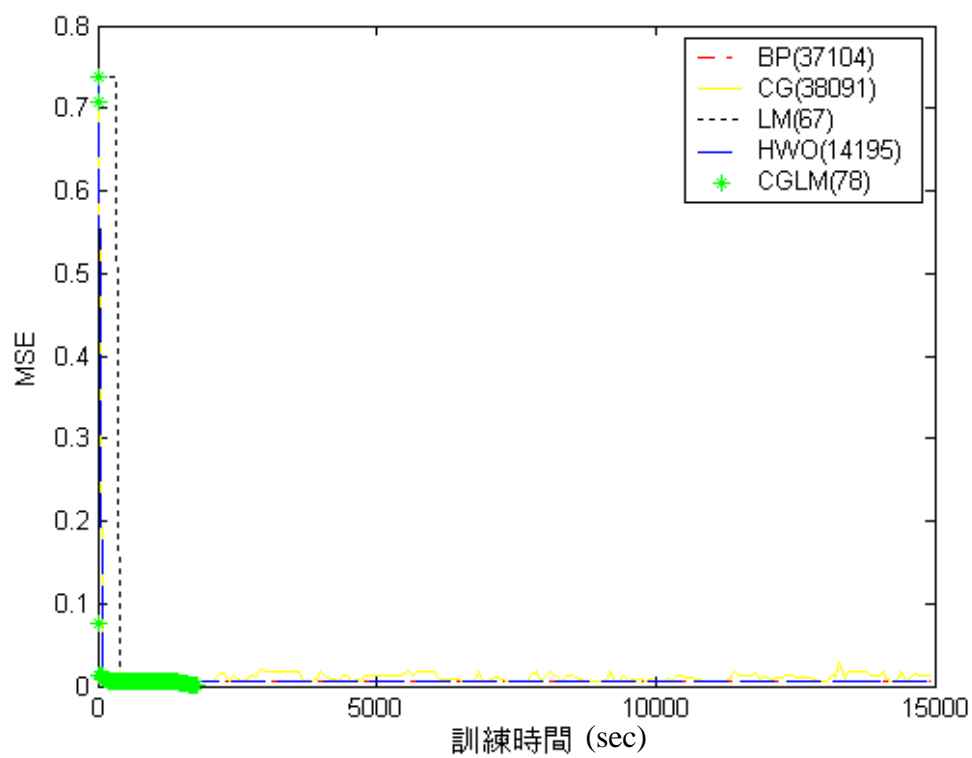


圖 4.60 類型三 (d) EX2 (2-40-1)

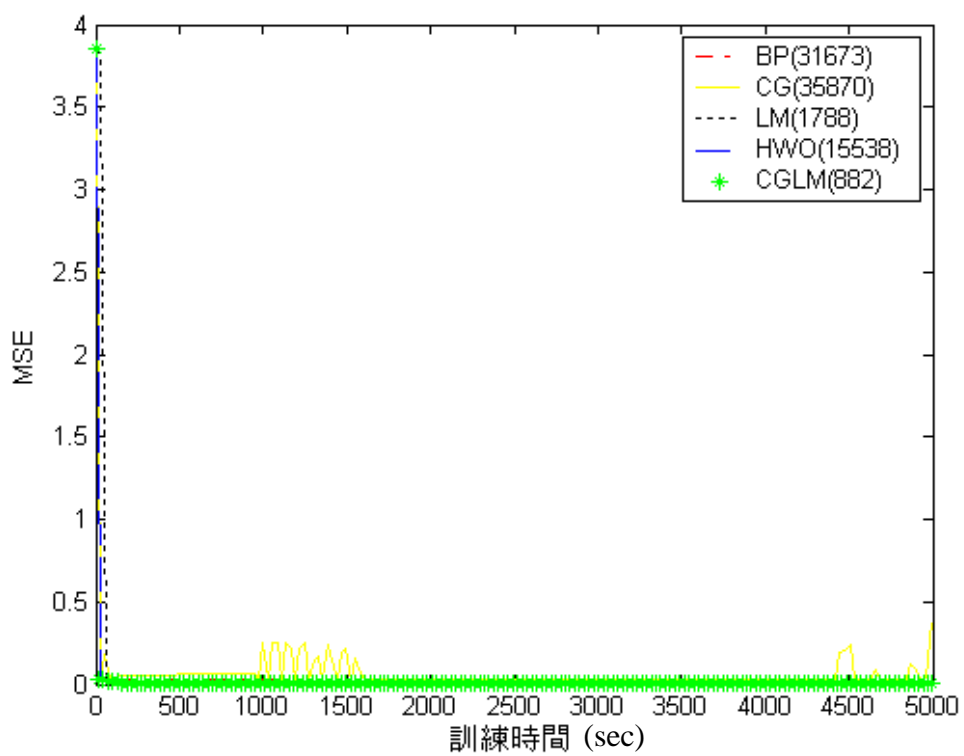


圖 4.61 類型三 (d) EX3 (2-20-1)

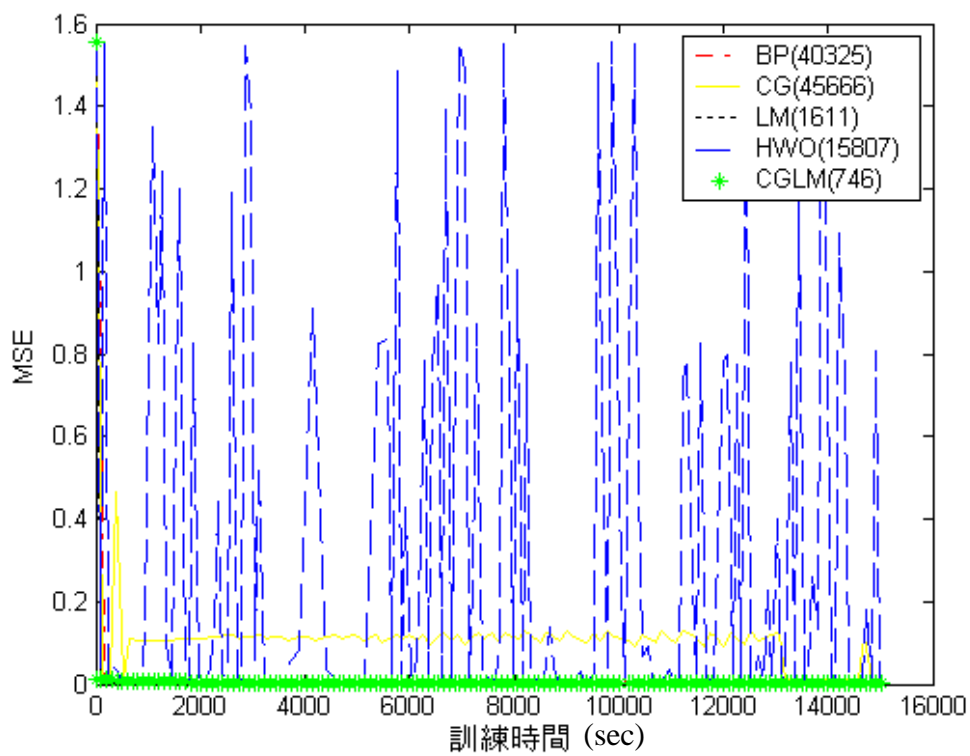


圖 4.62 類型三 (d) EX3 (2-50-1)

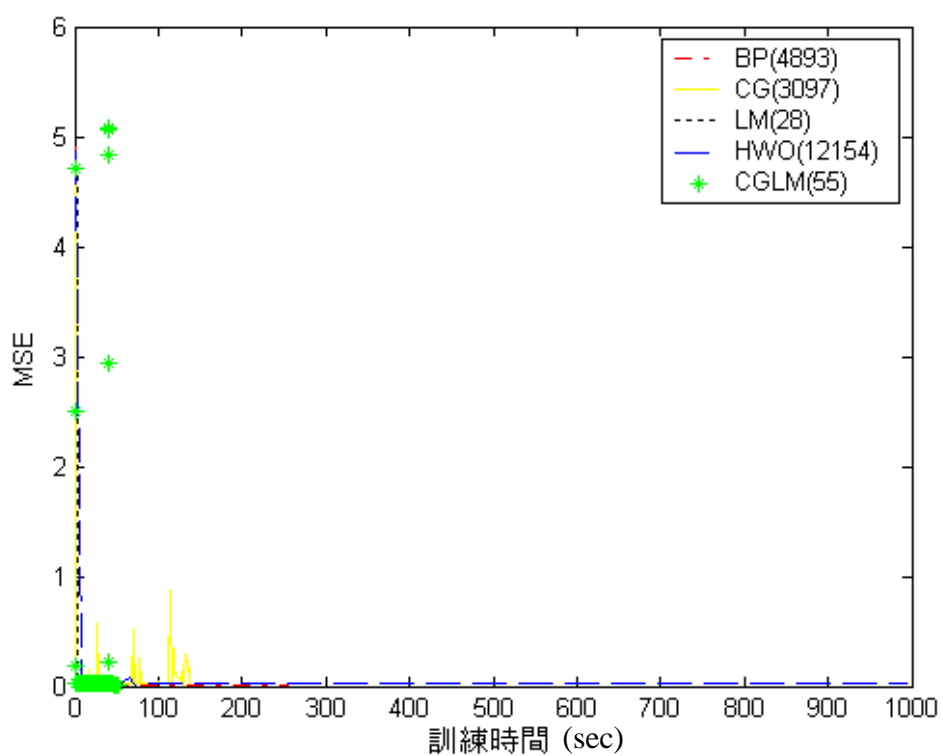


圖 4.63 類型三 (d) EX4 (2-8-1)

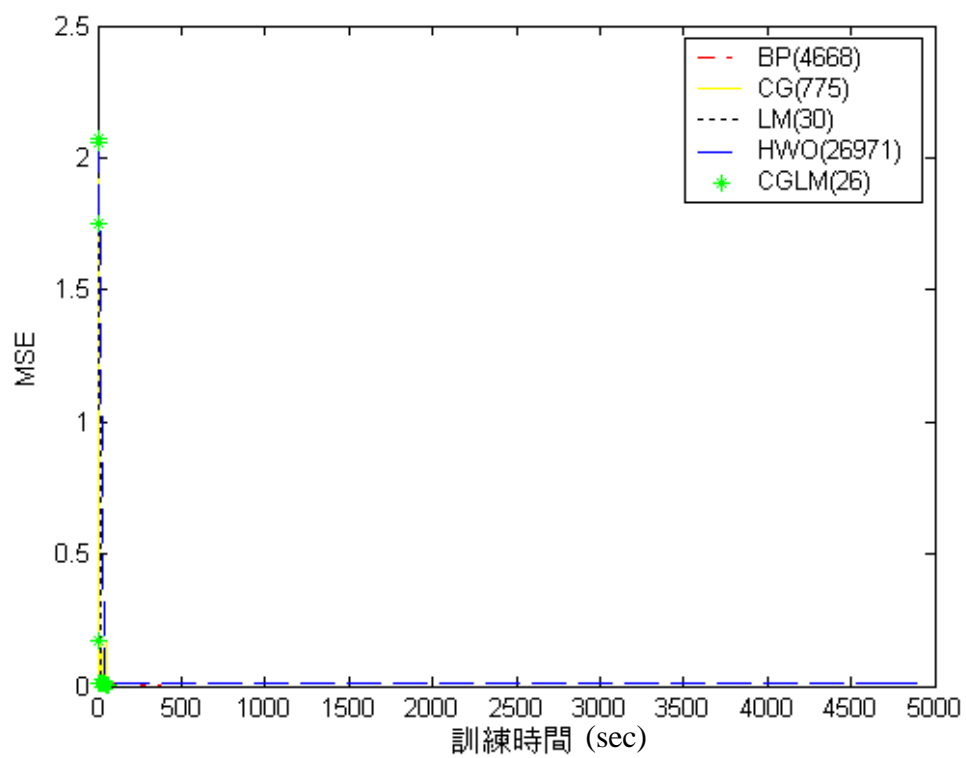


圖 4.64 類型三 (d) EX4 (2-20-1)

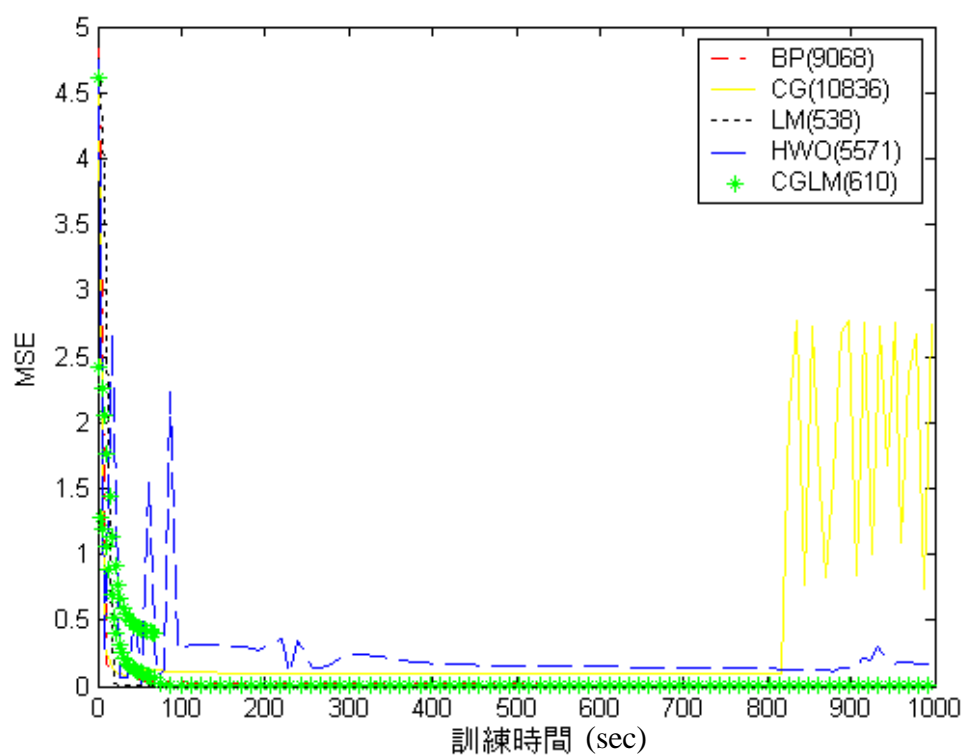


圖 4.65 類型三 EX5 (2-8-1)

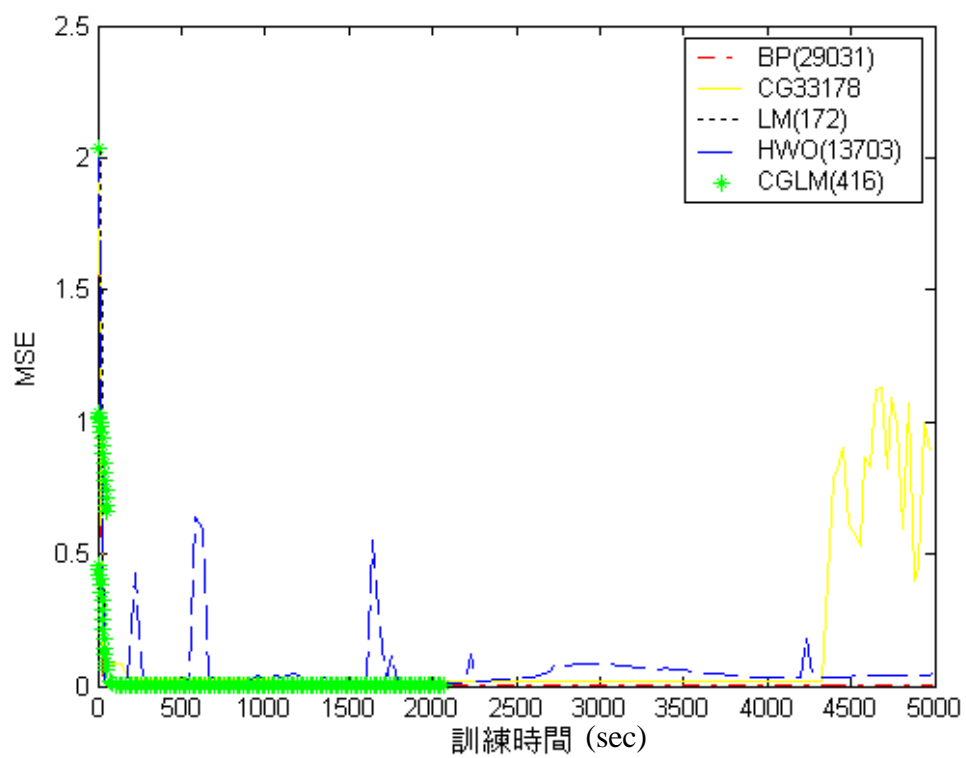


圖 4.66 類型三 (d) EX5 (2-20-1)

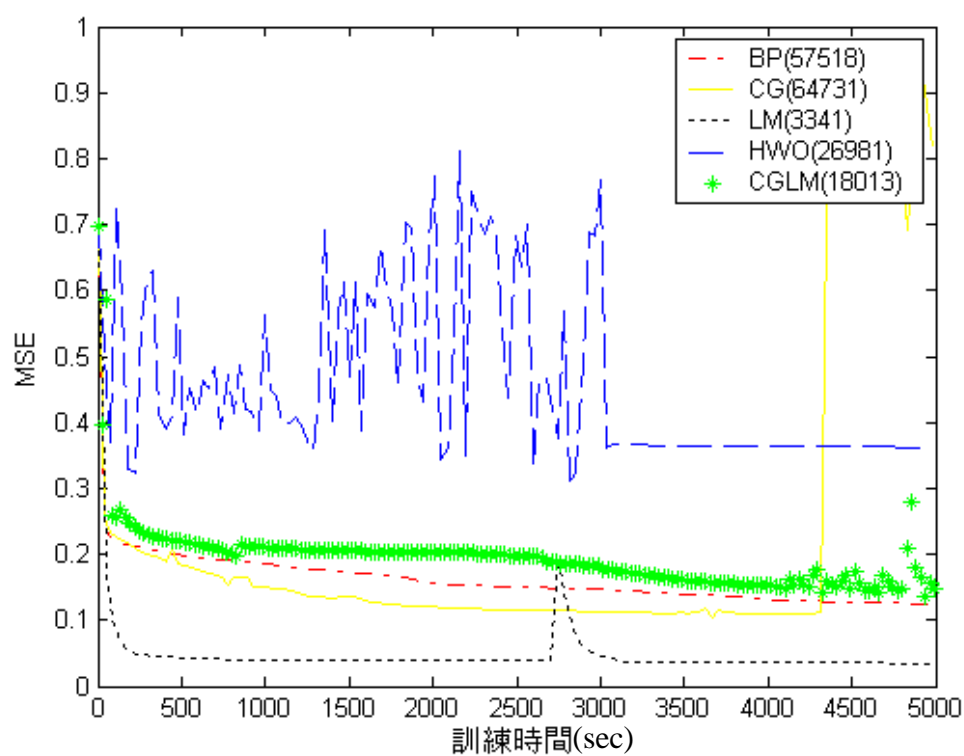


圖 4.67 類型三 (d) EX6 (2-20-1)

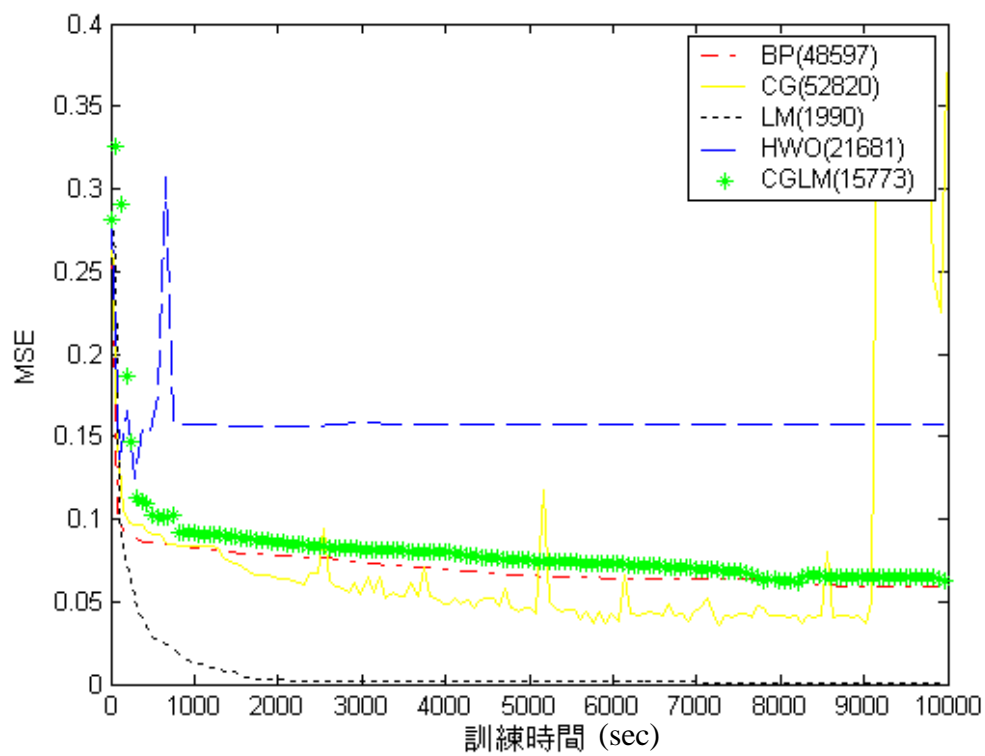


圖 4.68 類型三 (d) EX6 (2-50-1)

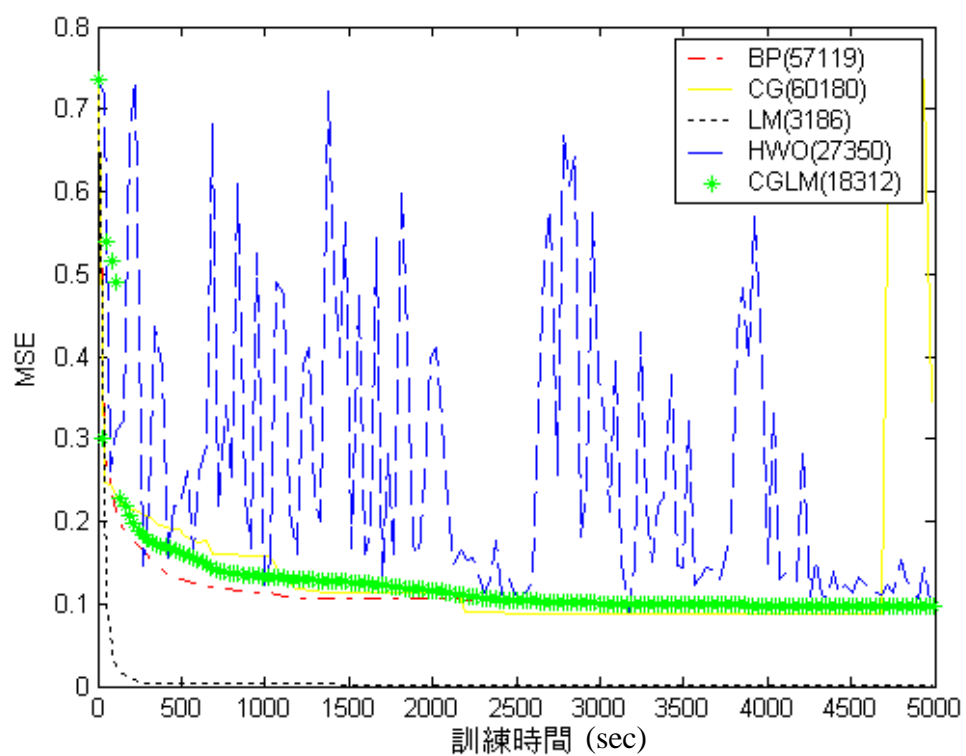


圖 4.69 類型三 (d) EX7 (2-20-1)

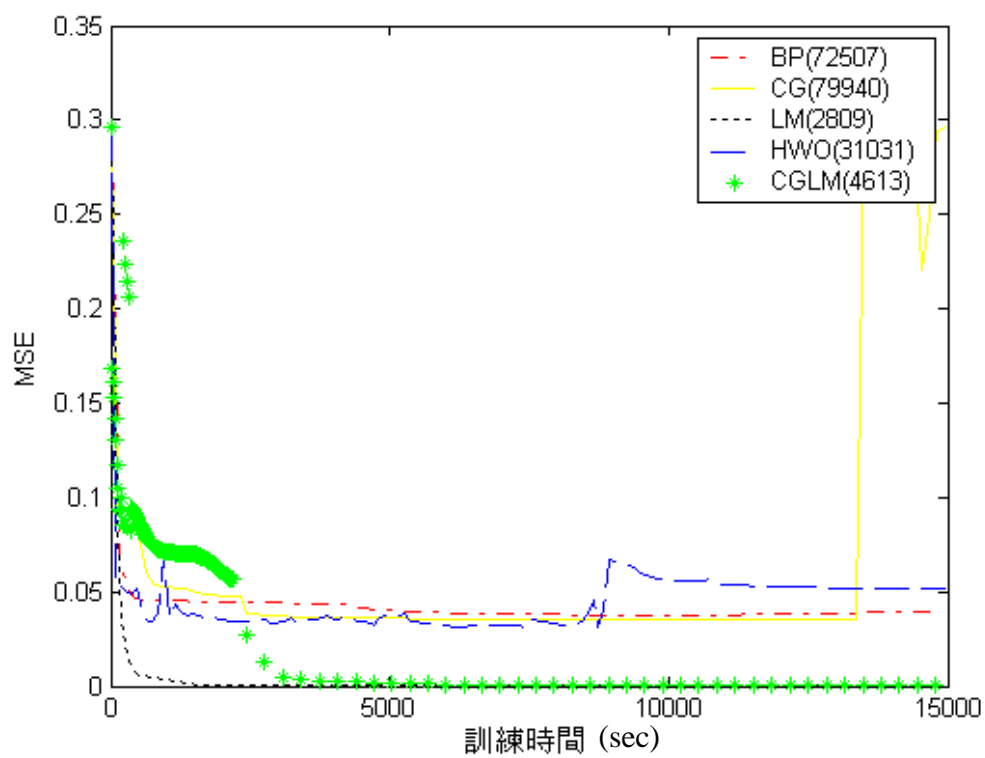


圖 4.70 類型三 (d) EX7 (2-50-1)

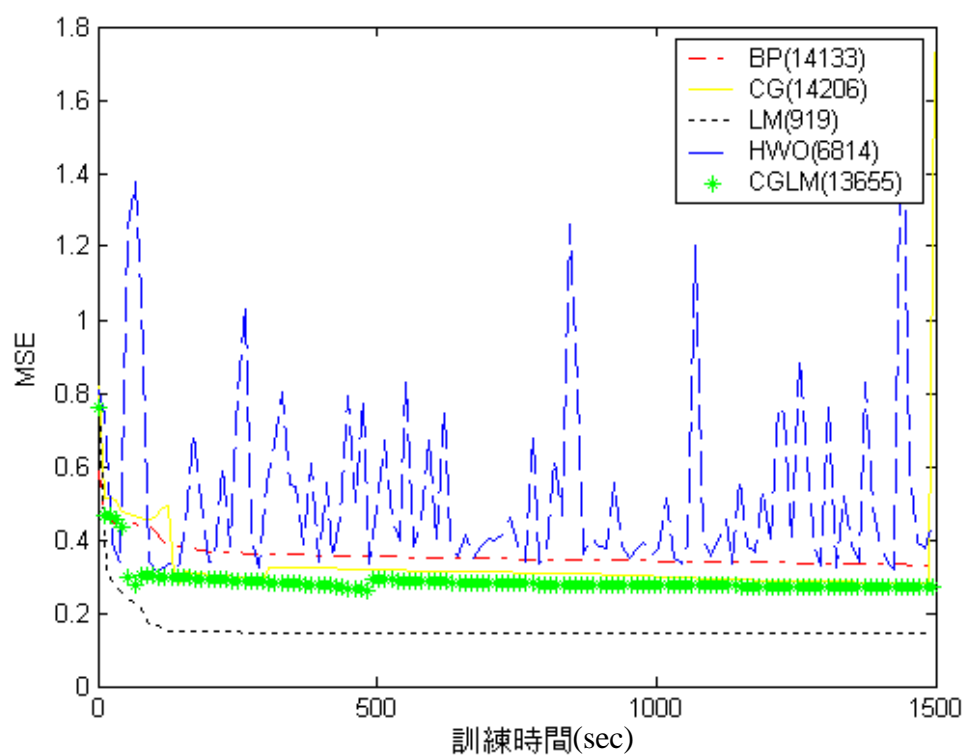


圖 4.71 類型三 (d) EX8 (2-8-1)

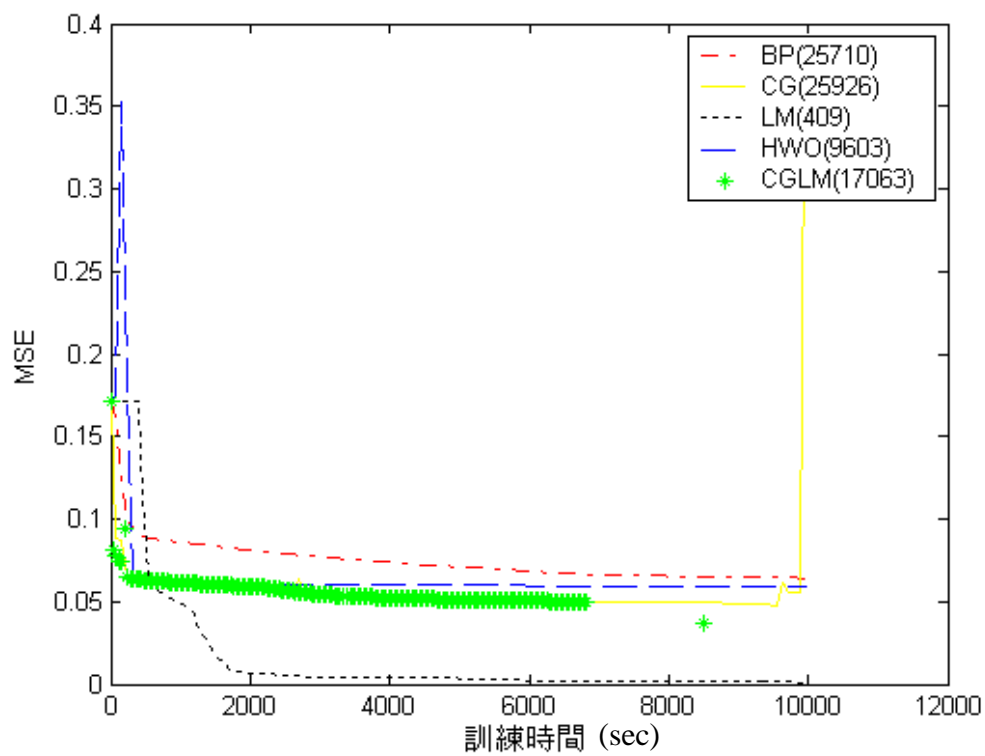


圖 4.72 類型三 (d) EX8 (2-40-1)

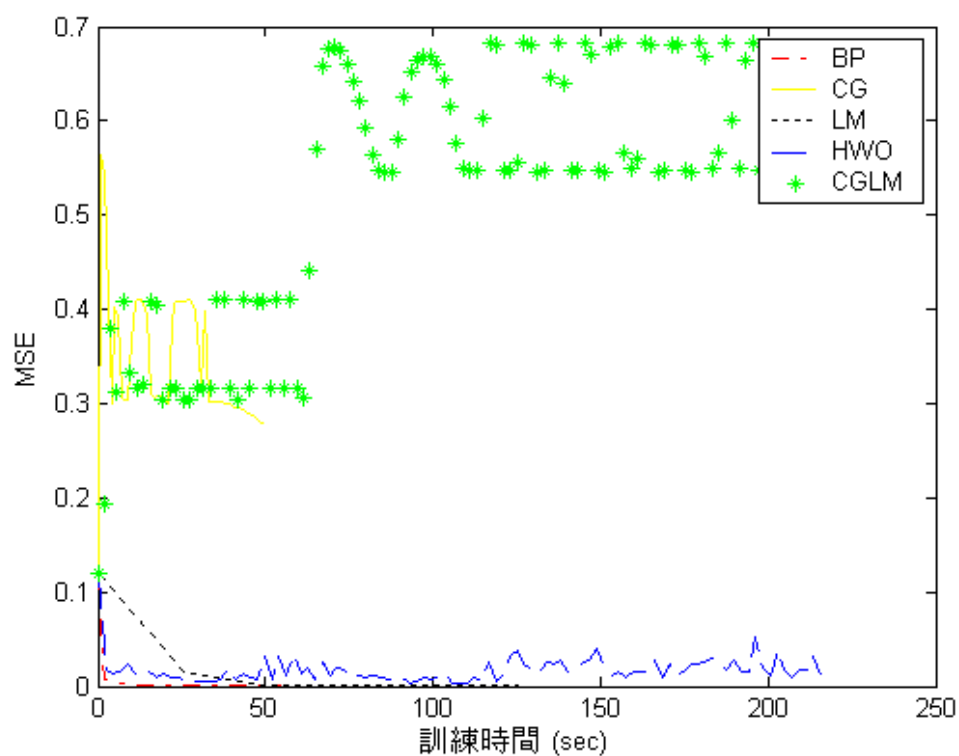


圖 4.73(a) 實例-展弦比等於 2 (5-12-5)

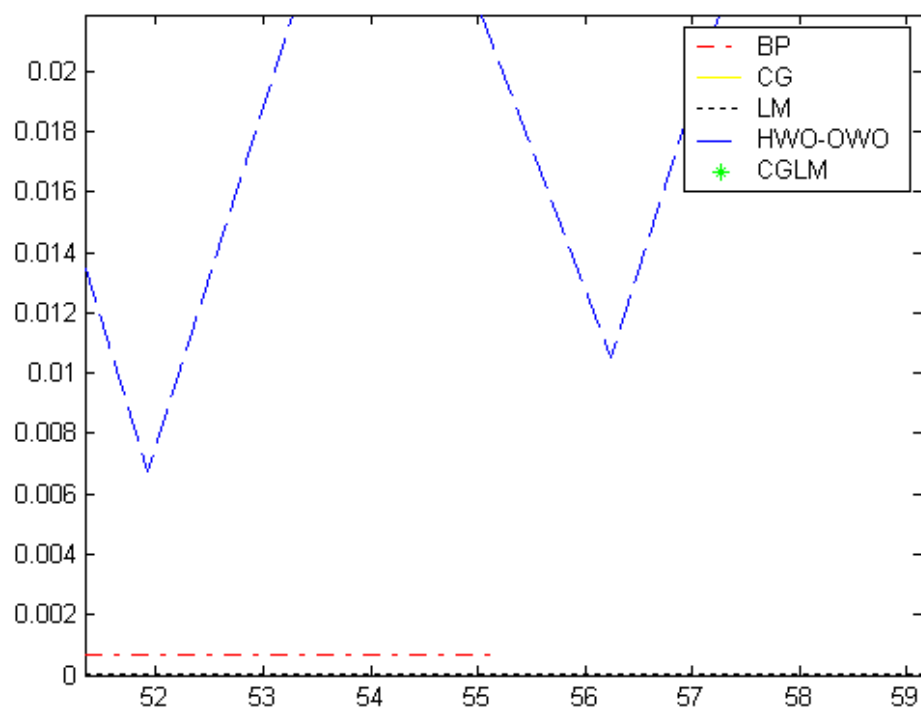


圖 4.73(b) 放大圖[圖 4.73(a)] (收斂誤差 10^{-5})

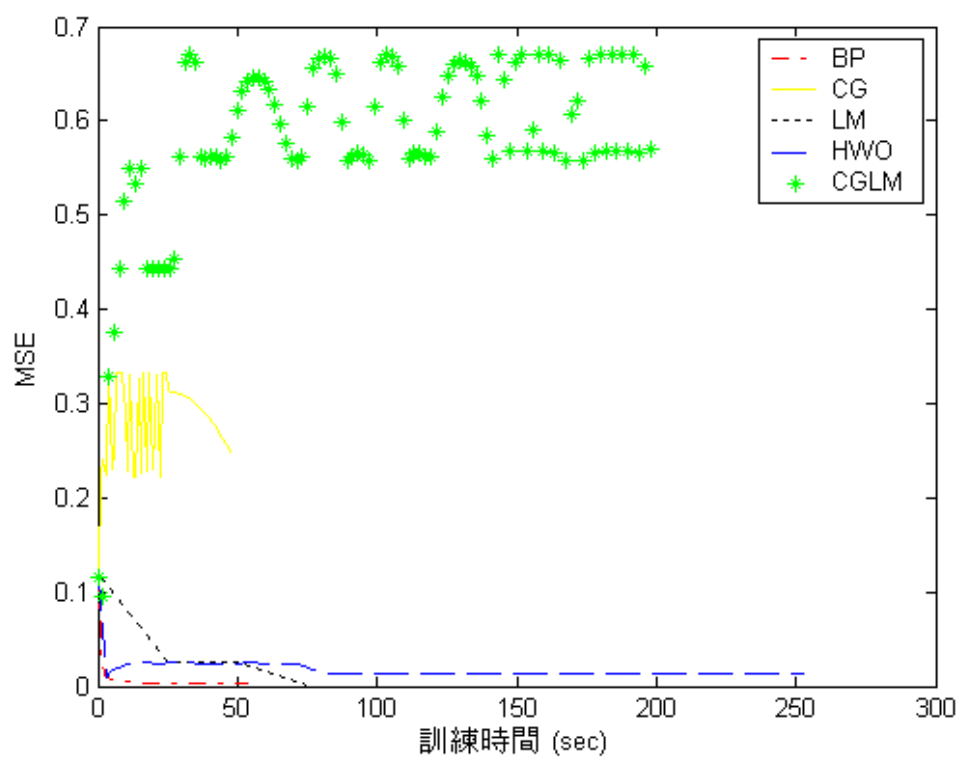


圖 4.74(a) 實例-展弦比等於 4 (5-12-5)

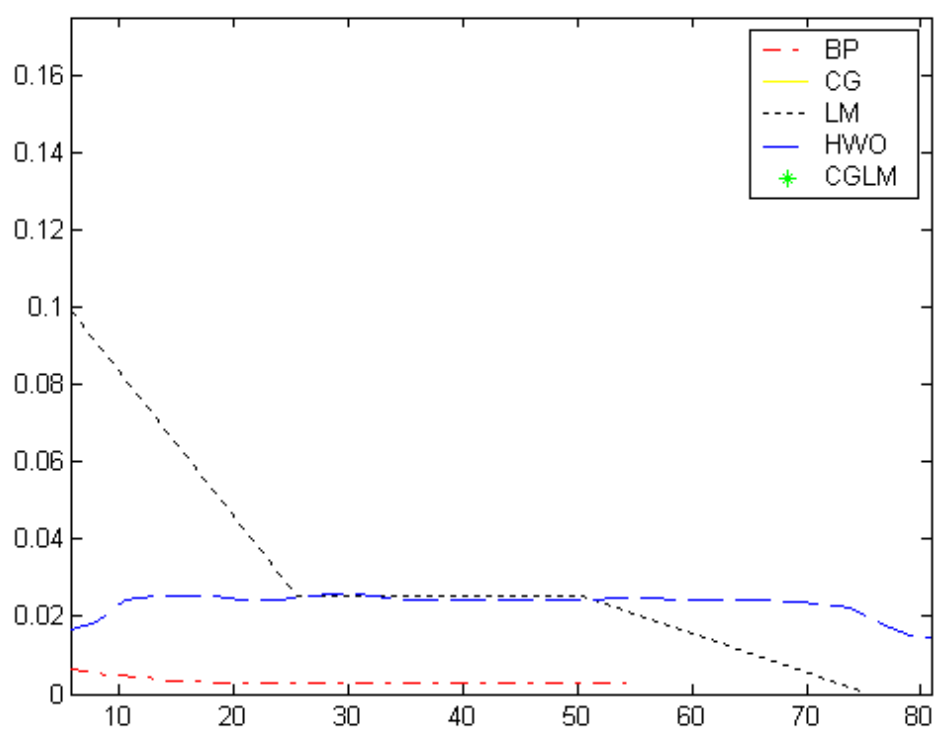


圖 4.74(b) 放大圖[圖 4.74(a)] (收斂誤差 10^{-5})

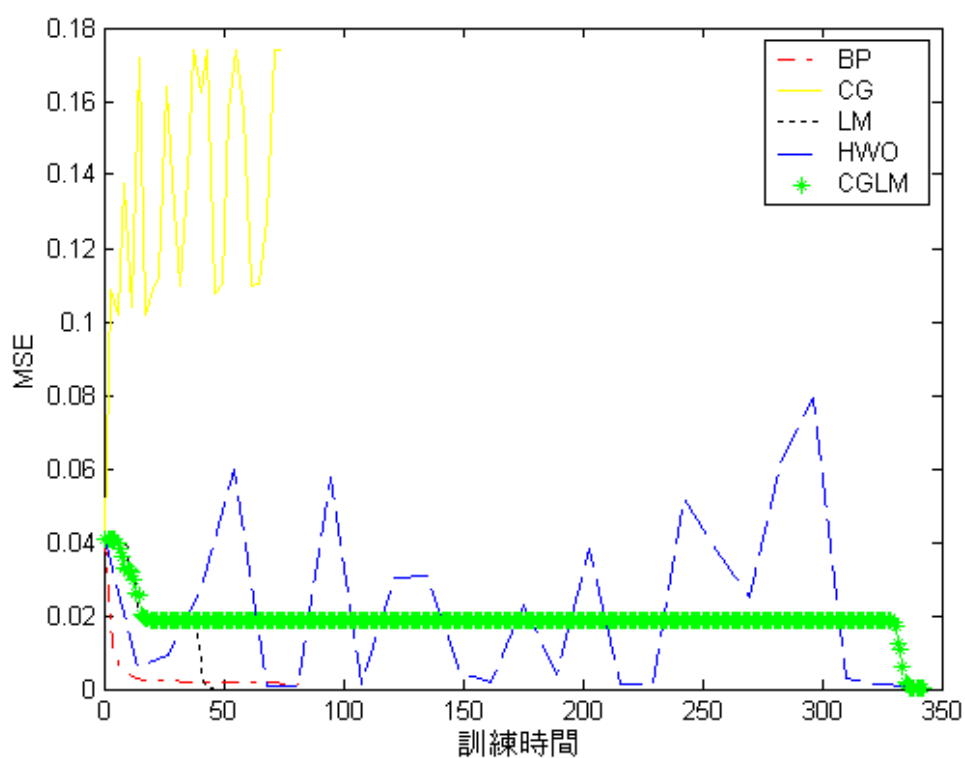


圖 4.75 實例-展弦比等於 2 (5-25-5)

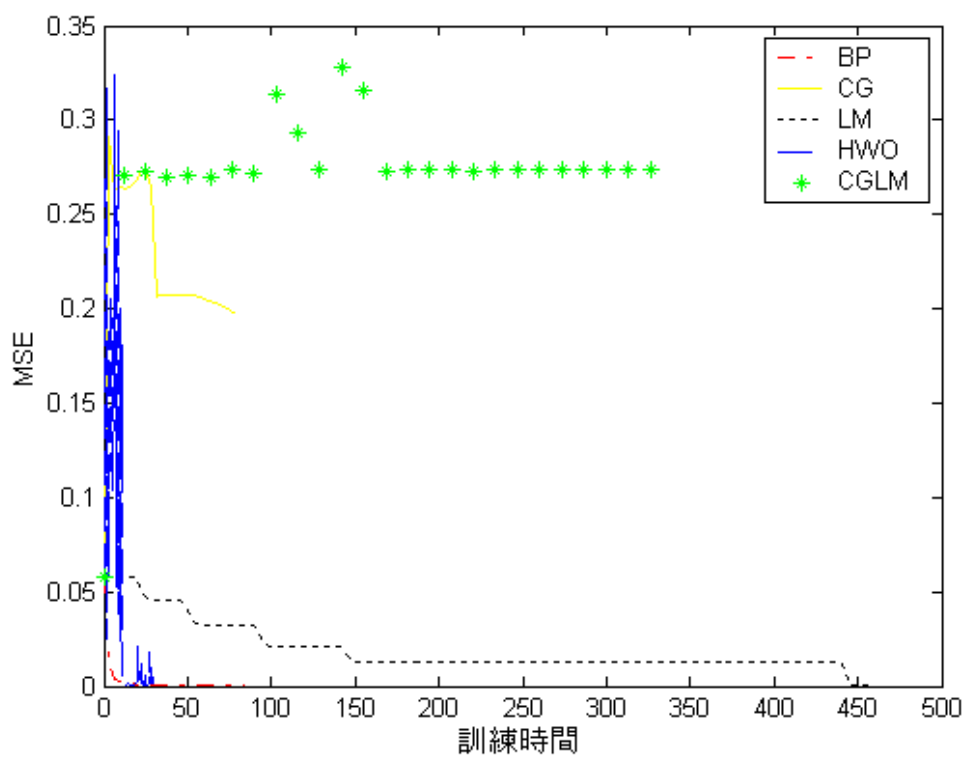


圖 4.76 實例-展弦比等於 4 (5-25-5)

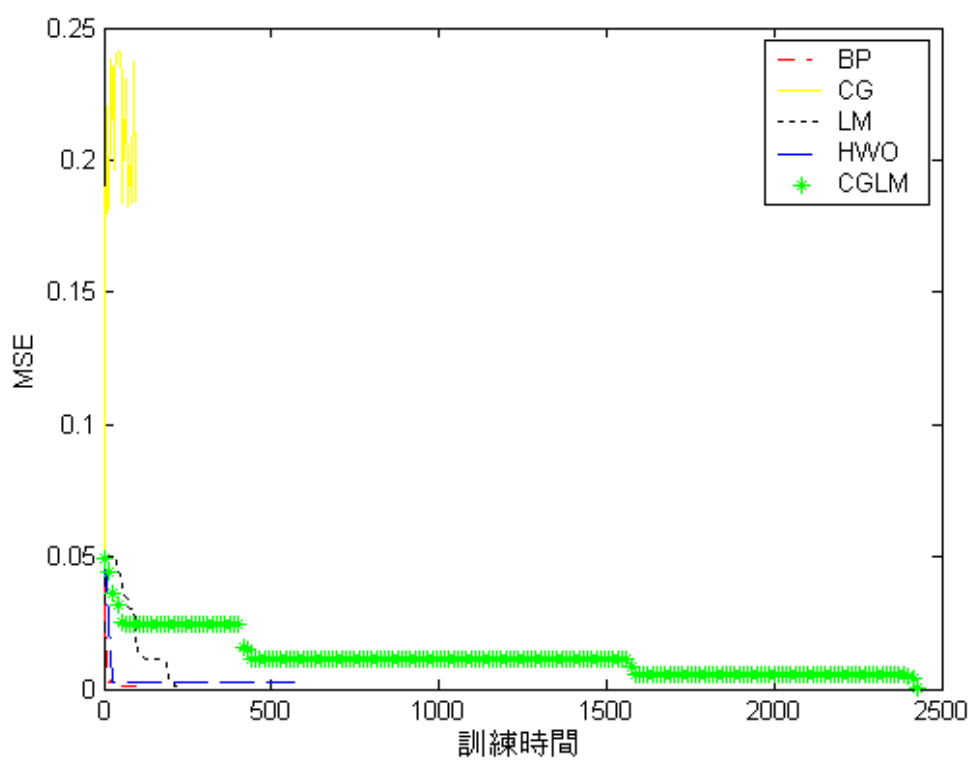


圖 4.77 實例-展弦比等於 2 (5-30-5)

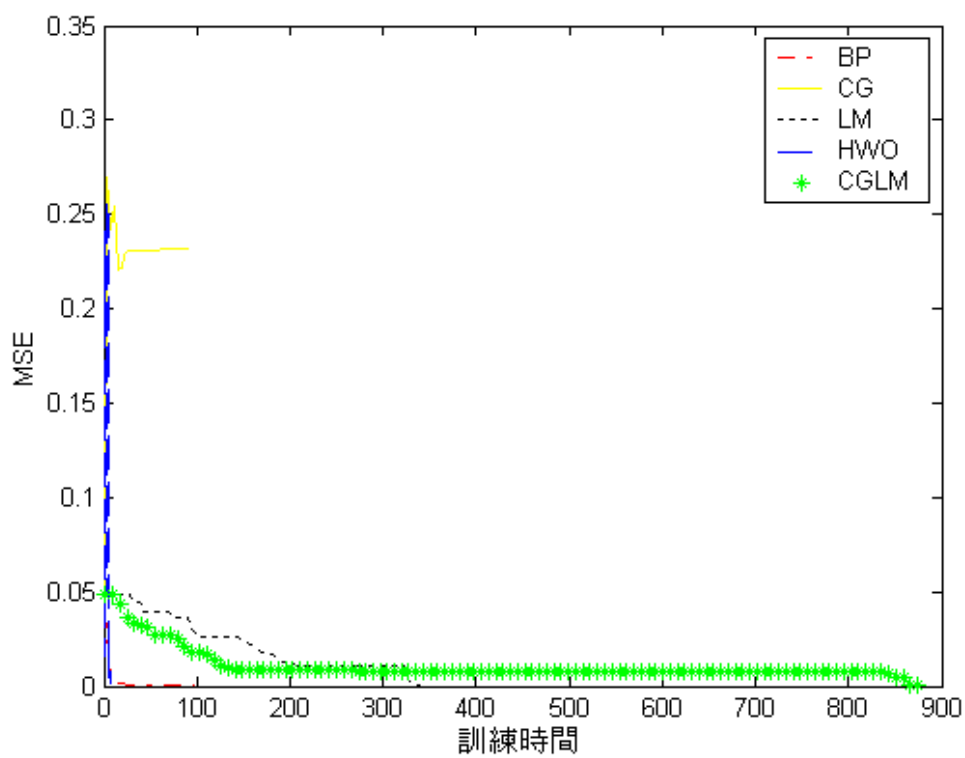


圖 4.78 實例-展弦比等於 4 (5-30-5)