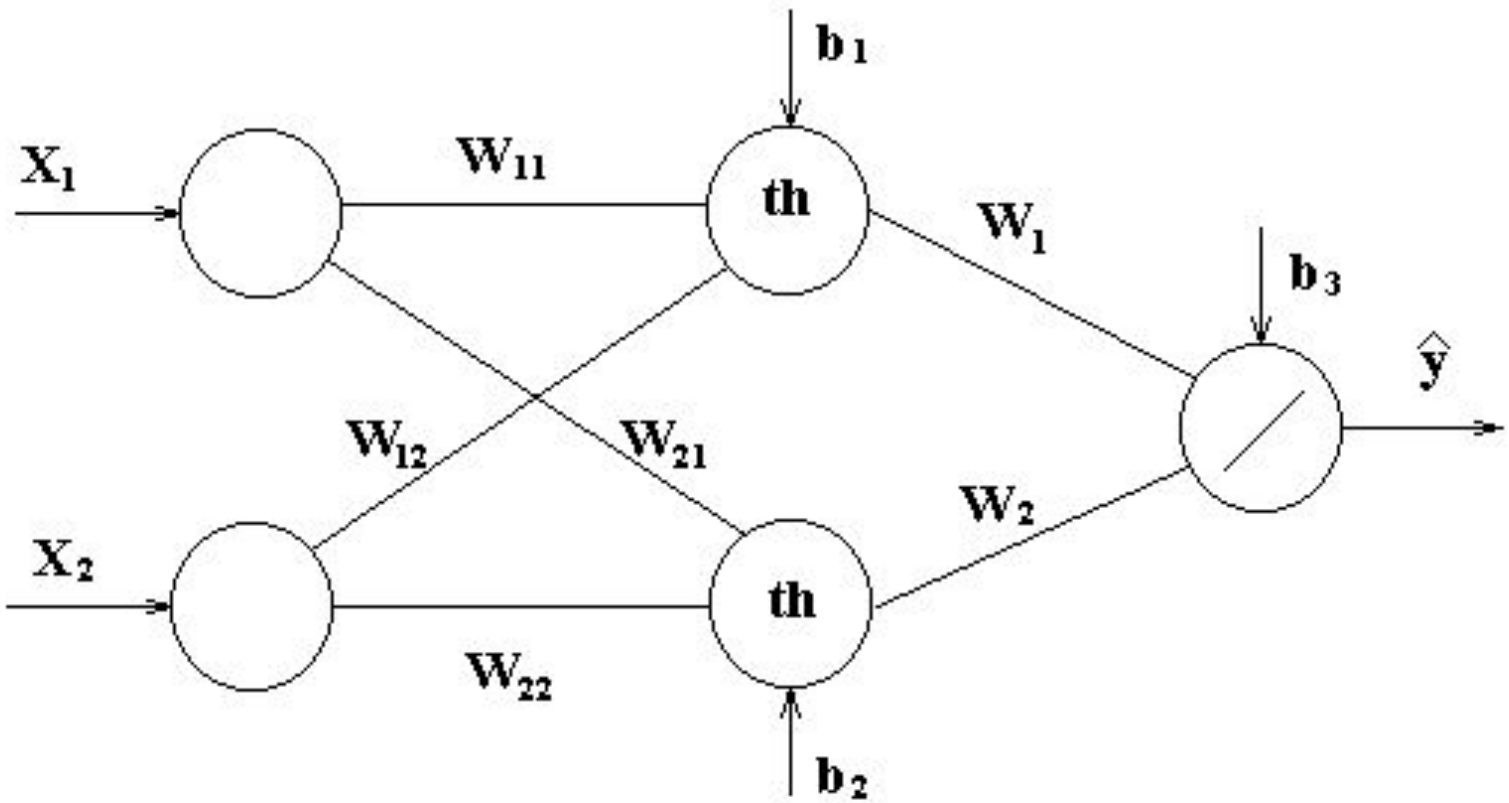


b 1
X1
W11
th
W1
b 3
ŷ
W12
W21
X2
W2
th
W22
b 2

Тогда для данной сети можно записать:

$$\hat{y}(t) = w_1 th(x_1 w_{11} + x_2 w_{12} + b_1) + w_2 th(x_1 w_{21} + x_2 w_{22} + b_2) + b_3.$$

Общий вид алгоритма:

$$\hat{\theta}(t) = \hat{\theta}(t-1) + K(t)e;$$

$$K(t) = Q(t)\psi(t);$$

$$Q(t) = \frac{P(t-1)}{\lambda + \psi(t)^T P(t-1)\psi(t)};$$

$$P(t) = \frac{1}{\lambda}\left(P(t-1) - \frac{P(t-1)\psi(t)\psi(t)^T P(t-1)}{\lambda + \psi(t)^T P(t-1)\psi(t)}\right).$$

Где $e = y(t) - \hat{y}(t)$ – ошибка сети;

$$\psi(t) = \begin{bmatrix} \frac{\delta\hat{y}}{\delta w_{11}} & \frac{\delta\hat{y}}{\delta w_{12}} & \frac{\delta\hat{y}}{\delta b_1} & \frac{\delta\hat{y}}{\delta w_{21}} & \frac{\delta\hat{y}}{\delta w_{22}} & \frac{\delta\hat{y}}{\delta b_2} & \frac{\delta\hat{y}}{\delta w_1} & \frac{\delta\hat{y}}{\delta w_2} & \frac{\delta\hat{y}}{\delta b_3} \end{bmatrix}^T.$$

Т.е. $\psi(t)$ – матрица Якоби, вычисленная с учетом текущих параметров сети.

Начальные условия:

$P(0) = \lambda^{-1} I$ ($\lambda = 0.99$, I – единичная матрица (9x9));

$\hat{\theta}(0)$ – вектор-столбец (9x1) малых случайных величин (0.001).

Алгоритм обратного распространения ошибки

Для реализации алгоритма обратного распространения необходимо:

1. Выбрать из заданного обучающего множества очередную обучающую пару $(x(i), y^{*}(i))$, $i = \overline{1,P}$, и подать на вход сети входной сигнал $x(i)$.
2. Вычислить реакцию сети $y(i)$.
3. Сравнить полученную реакцию $y(i)$ с требуемой $y^{*}(i)$ и определить ошибку $y^{*}(i) - y(i)$.
4. Скорректировать веса так, чтобы ошибка была минимальной.
5. Шаги 1-4 повторить для всего множества обучающих пар $(x(i), y^{*}(i))$ $i = \overline{1,P}$ до тех пор, пока на заданном множестве ошибка не достигнет требуемой величены.

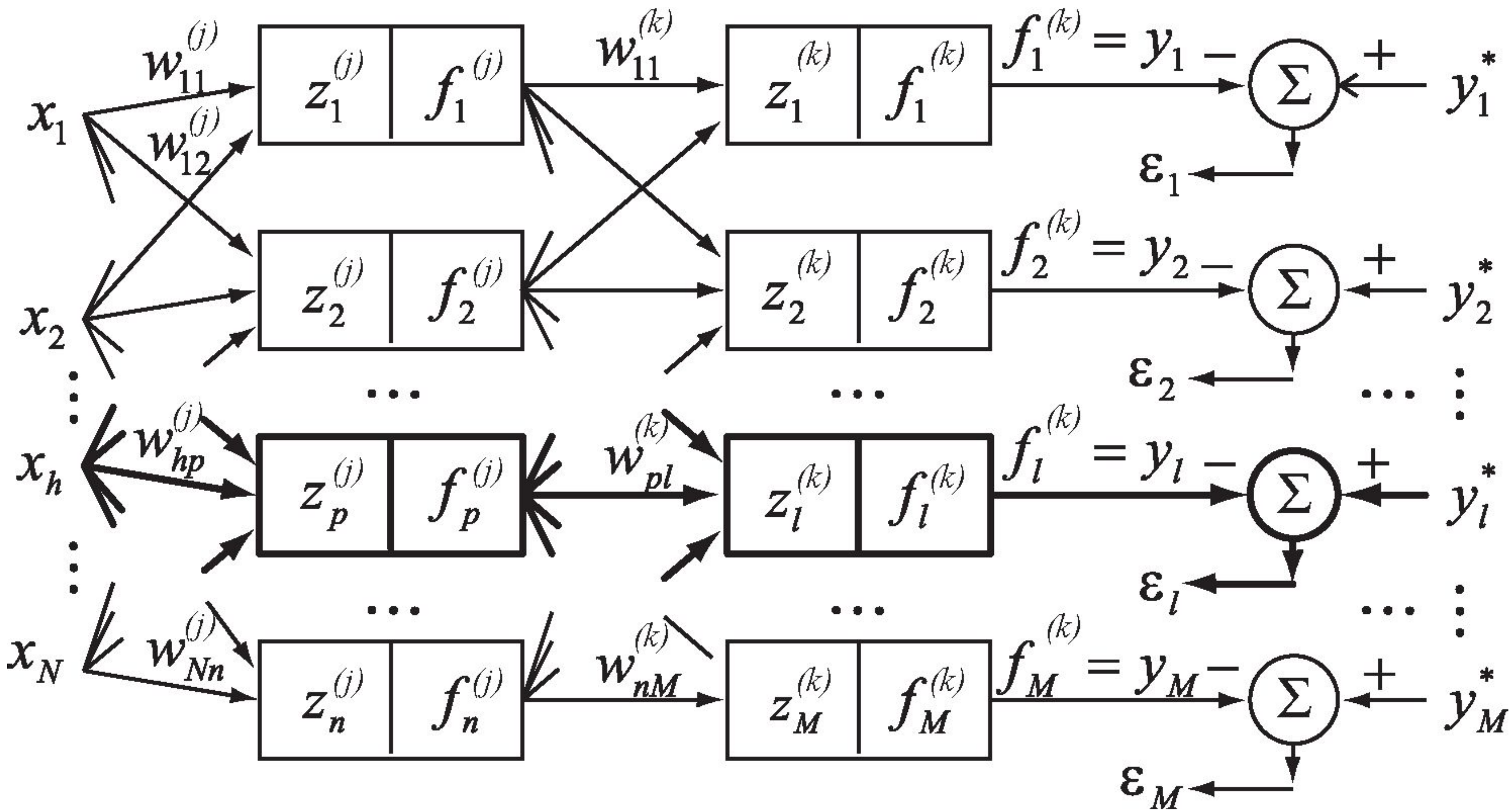


Рис.3.25 – Фрагмент сети прямого распространения

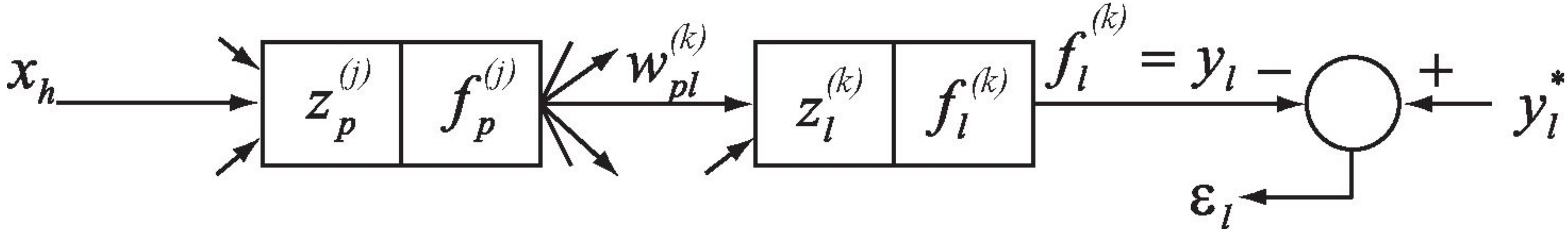


Рис.3.26 – Фрагмент многослойной сети

Использование квадратичного критерия качества обучения

$$\zeta_l^2 = (y_l^* - f_l^{(k)})^2 \qquad (3.35)$$

позволяет получить градиентный алгоритм коррекции весов, который в данном случае принимает вид

$$\Delta W_{pl}^{(k)} = -\gamma_{pl} \frac{\partial \varsigma_l^2}{\partial W_{pl}^{(k)}}, \qquad (3.36)$$

где γ_{pl} - коэффициент, влияющий на скорость обучения.

Вычисление входящей в (3.36) частной производной осуществляется по правилу

$$\frac{\partial \varsigma_l^2}{\partial W_{pl}^{(k)}} = \frac{\partial \varsigma_l^2}{\partial f_l^{(k)}} \frac{\partial f_l^{(k)}}{\partial Z_l^{(k)}} \frac{\partial Z_l^{(k)}}{\partial W_{pl}^{(k)}}. \qquad (3.37)$$

С учетом (3.35), вида активационных функций и того, что

$$Z_l^{(k)} = \sum_{p=1}^{n} W_{pl}^{(k)} f_p^{(j)},$$

получаем

$$\frac{\partial \zeta_l^2}{\partial f_l^{(k)}} = -2(y_l^* - y_l) = -2\zeta_l; \tag{3.38}$$

$$\frac{\partial f_l^{(k)}}{\partial Z_l^{(k)}} = \alpha f_l^{(k)}(1 - f_l^{(k)}); \tag{3.39}$$

$$\frac{\partial Z_l^{(k)}}{\partial W_{pl}^{(k)}} = f_p^{(k)}. \tag{3.40}$$

Подстановка (3.38) – (3.40) в (3.37) дает

$$\Delta W_{pl}^{(k)} = 2\alpha\gamma_{pl}\zeta_l f_l^{(k)}(1 - f_l^{(k)}) f_p^{(k)} \tag{3.41}$$

или

$$W_{pl}^{(k)}(i+1) = W_{pl}^{(k)}(i) + \gamma_{pl}\delta_{pl}^{(k)} f_p^{(k)}, \tag{3.42}$$

где

$$\delta_{pl}^{(k)} = 2\alpha\zeta_l f_l^{(k)}(1 - f_l^{(k)}), \tag{3.43}$$

i – номер итерации.

Вычисление весов нейронов скрытого слоя.

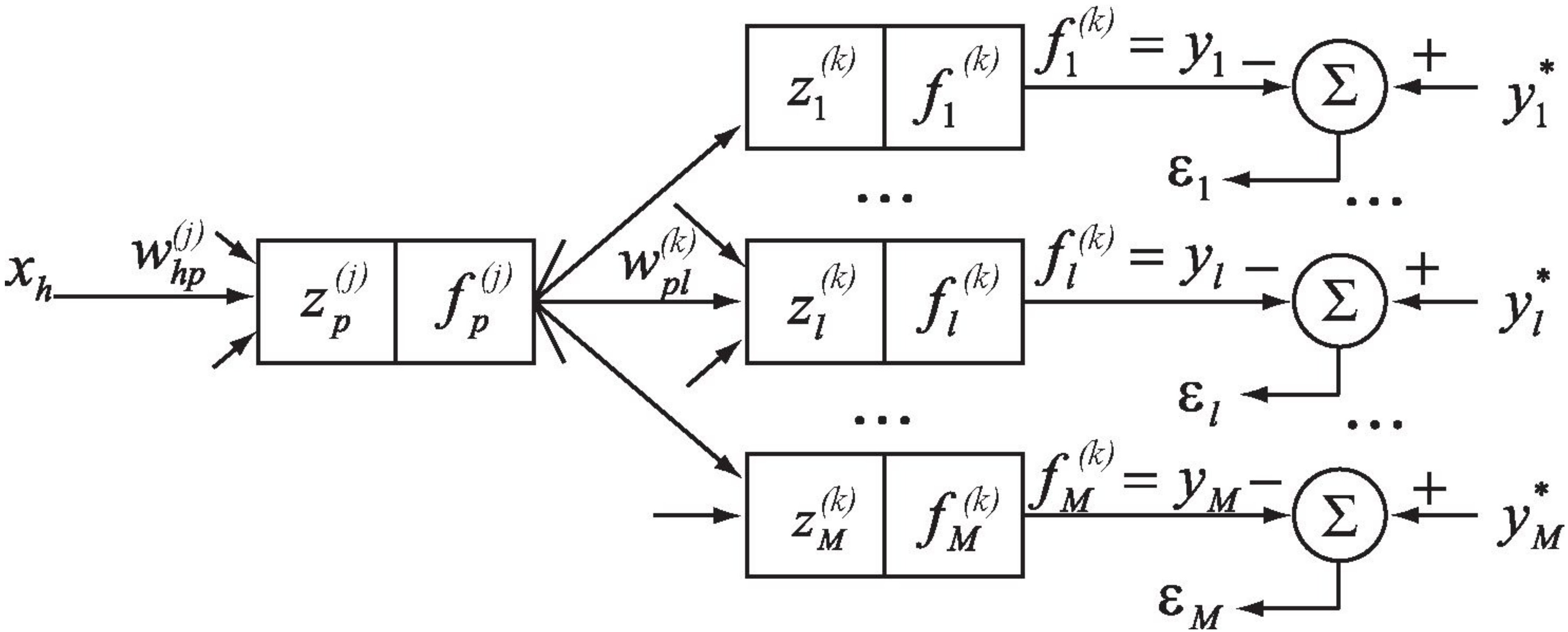


Рис.3.27 – Фрагмент сети

Рассмотрим р- й нейрон скрытого (j-го) слоя. Для вычисления веса $w_{np}^{(j)}$ данного нейрона также используется дельта-правило, аналогичное (3.35). Но так как выходной сигнал этого нейрона поступает на выходы всех нейронов входного слоя, алгоритм настройки записывается следующим образом:

$$\Delta W_{hp}^{(j)} = -\gamma_{hp} \frac{\partial \zeta^2}{\partial W_{hp}^{(j)}} = -\gamma_{hp} \sum_{i-1}^{M} \frac{\partial \zeta_l^{(2)}}{\partial W_{hp}^{(j)}}. \tag{3.44}$$

Здесь ζ_i - ошибка на i-м выходе; γ_{hp} - параметр, играющий ту же роль, что и γ_{pl} в (3.42).

Так как квадратичный критерий качества в этом случае принимает вид

$$\zeta = \sum_{l-1}^{M} \left[y_l^* - f_l^{(k)} \right]^2, \tag{3.45}$$

частная производная, используемая в (3.44), вычисляется так

$$\frac{\partial \zeta^2}{\partial W_{hp}^{(j)}} = \sum_{l=1}^{M} \frac{\partial \zeta_l^2}{\partial f_l^{(k)}} \frac{\partial f_l^{(k)}}{\partial Z_l^{(k)}} \frac{\partial Z_l^{(k)}}{\partial f_p^{(j)}} \frac{\partial f_p^{(j)}}{\partial Z_p^{(j)}} \frac{\partial Z_p^{(j)}}{\partial W_{hp}^{(j)}}. \tag{3.46}$$

Для рассматриваемого случая по аналогии с вышеизложенным получаем

$$\frac{\partial \zeta_l^2}{\partial f_l^{(k)}} = -2[y_l^* - f_l^{(k)}] = -2\zeta_l; \tag{3.47}$$

$$\frac{\partial f_i^{(k)}}{\partial Z_i^{(k)}} = \alpha f_i^{(k)}(1 - f_i^{(k)}); \tag{3.48}$$

$$\frac{\partial Z_l^{(k)}}{\partial f_p^{j}} = W_{pl}^{(k)}; \tag{3.49}$$

$$\frac{\partial f_p^{(j)}}{\partial Z_p^{(j)}} = \alpha f_p^{(j)}(1 - f_p^{(j)}); \tag{3.50}$$

$$\frac{\partial Z_p^{(j)}}{\partial W_{hp}^{(j)}} = x_h. \tag{3.51}$$

Здесь учтено, что

$$Z_l^{(k)} = \sum_{p=1}^{m} W_{pl}^{(k)} f_p^{(j)};$$

$$Z_l^{(j)} = \sum_{h=1}^{N} W_{hp}^{(j)} x_h,$$

а функции активации являются сигмоидальными.

Таким образом,

$$\frac{\partial\zeta^2}{\partial W_{hp}^{(j)}}=\sum_{l=1}^{M}(-2)\alpha\zeta_l\left[f_l^{(k)}(1-f_l^{(k)})\right]W_{pl}^{(k)}\alpha\left[f_p^{(j)}(1-f_p^{(j)})\right]x_h=-\sum_{l=1}^{M}\delta_{pl}^{(k)}W_{pl}^{(k)}\frac{\partial f_p^{(j)}}{\partial Z_p^{(j)}}x_h \tag{3.52}$$

где δ_{pl}^{k} определяется выражением (3.43).

Обозначим

$$\delta_{hp}^{(j)}=\delta_{pl}^{(k)}W_{pl}^{(k)}\frac{\partial f_p^{(j)}}{\partial Z_p^{(j)}}. \tag{3.53}$$

Тогда

$$\frac{\partial\zeta^2}{\partial W_{hp}^{(j)}}=-\sum_{l=1}^{M}\delta_{hp}^{(j)}x_h \tag{3.54}$$

и алгоритм коррекции весов принимает вид

$$\Delta W_{hp}^{(j)}=-\gamma_{hp}x_h\sum_{l=1}^{M}\delta_{hp}^{(j)} \tag{3.55}$$

или

$$W_{hp}^{(j)}(i+1)=W_{hp}^{(j)}(i+1)+\gamma_{hp}x_h\sum_{l=1}^{M}\delta_{hp}^{(j)}. \tag{3.56}$$

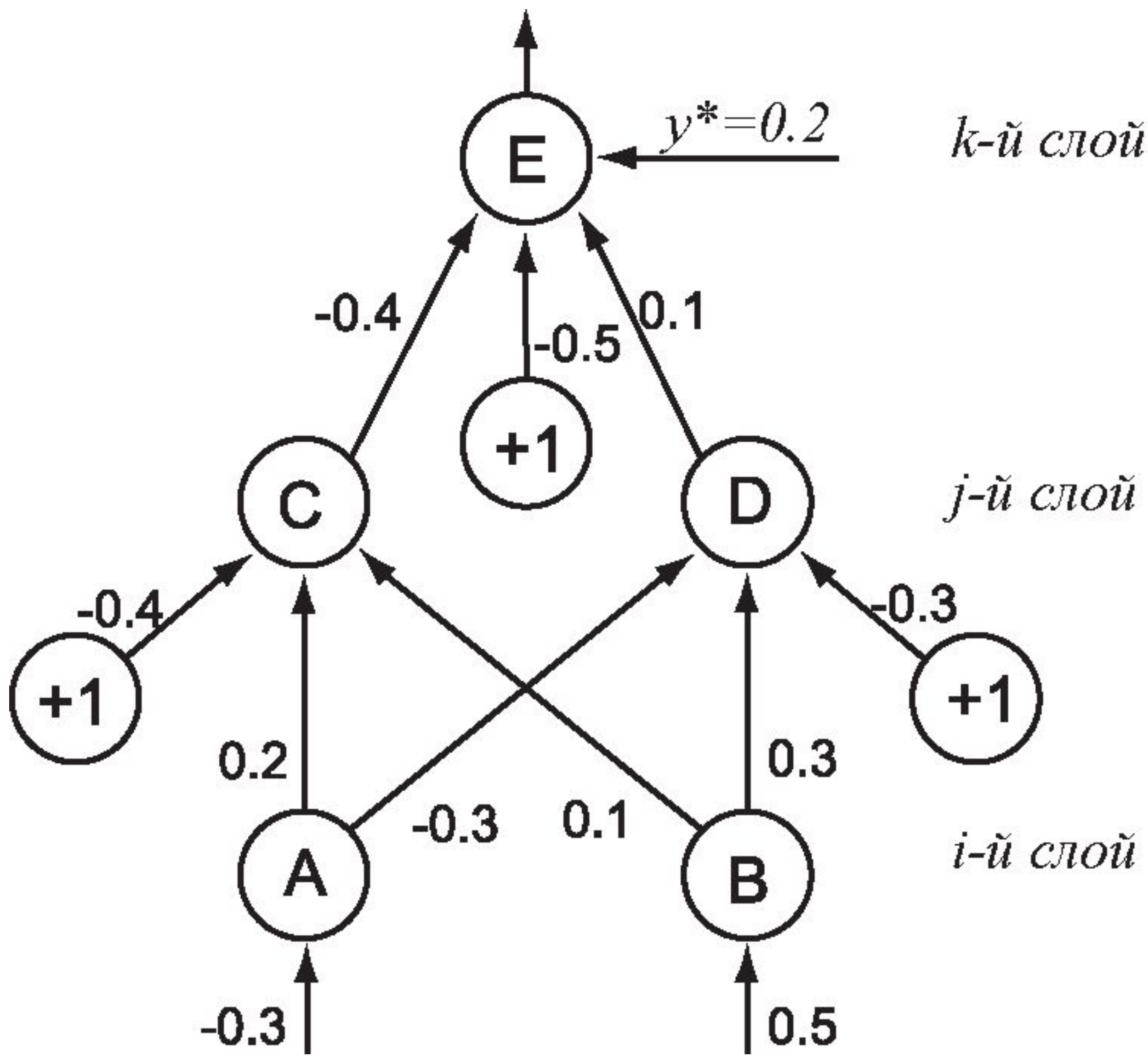


Рис.3.28 – Пример сети

Веса между слоями j и k корректируются по правилу (3.42), а между слоями i и j – по правилу (3.56).

Сначала вычислим выходы каждого нейрона, учитывая вид функции активации $f=(1+e^{-\alpha x})^{-1}$ при $\alpha=1$;

$Z_C=-0.3\cdot 0.2+0.5\cdot 0.1+1\cdot(-0.4)=-0.06+0.05-0.4=-0.41;$ $\quad y_C=f_C=0.399;$

$Z_D=-0.3\cdot(-0.3)+0.5\cdot 0.3+1\cdot(-0.3)=0.09+0.15-0.3=-0.06;$ $\quad y_D=f_D=0.485;$

$Z_E=0.399\cdot(-0.4)+0.485\cdot 0.1+1\cdot(-0.5)=-0.16+0.485-0.5=-0.17;$ $\quad y_E=f_E=0.352.$

Подстановка этих и других значений в выражения (3.41) и (3.55) дает

$\Delta W_{CE}=-0.5\cdot(-2)\cdot 1\cdot(0.2-0.352)\cdot 0.352\cdot(1-0.352)\cdot 0.399=-0.01381;$

$\Delta W_{DE}=(0.2-0.352)\cdot 0.352\cdot(1-0.352)\cdot 0.485=-0.01679;$

$\Delta W_{1E}=(0.2-0.352)\cdot 0.352\cdot(1-0.352)\cdot 1=-0.03462;$

$\Delta W_{AC}=(0.2-0.352)\cdot 0.352\cdot(1-0.352)\cdot(-0.4)\cdot 1\cdot 0.399\cdot(1-0.399)\cdot(-0.3)=-0.001;$

$\Delta W_{AD}=(0.2-0.352)\cdot 0.352\cdot(1-0.352)\cdot 0.1\cdot 1\cdot 0.485\cdot(1-0.485)\cdot(-0.3)=0.00026;$

$\Delta W_{BC}=(0.2-0.352)\cdot 0.352\cdot(1-0.352)\cdot(-0.4)\cdot 1\cdot 0.399\cdot(1-0.399)\cdot 0.5=0.00166;$

$\Delta W_{BD}=(0.2-0.352)\cdot 0.352\cdot(1-0.352)\cdot 0.1\cdot 1\cdot 0.485\cdot(1-0.485)\cdot 0.5=-0.00043;$

$\Delta W_{1C}=(0.2-0.352)\cdot 0.352\cdot(1-0.352)\cdot(-0.4)\cdot 1\cdot 0.399\cdot(1-0.399)\cdot 1=0.00332;$

$\Delta W_{AD}=(0.2-0.352)\cdot 0.352\cdot(1-0.352)\cdot 0.1\cdot 1\cdot 0.485\cdot(1-0.485)\cdot 1=-0.00086.$

Добавив эти приращения к существующим весам, т.е. воспользовавшись формулами (3.42) и (3.55), получаем следующие новые значения весов:

$W_{CE} = -0.4 - 0.01381 = -0.414;$

$W_{DE} = 0.1 - 0.01679 = 0.083;$

$W_{1E} = -0.5 - 0.03462 = -0.535;$

$W_{AC} = 0.2 - 0.001 = 0.199;$

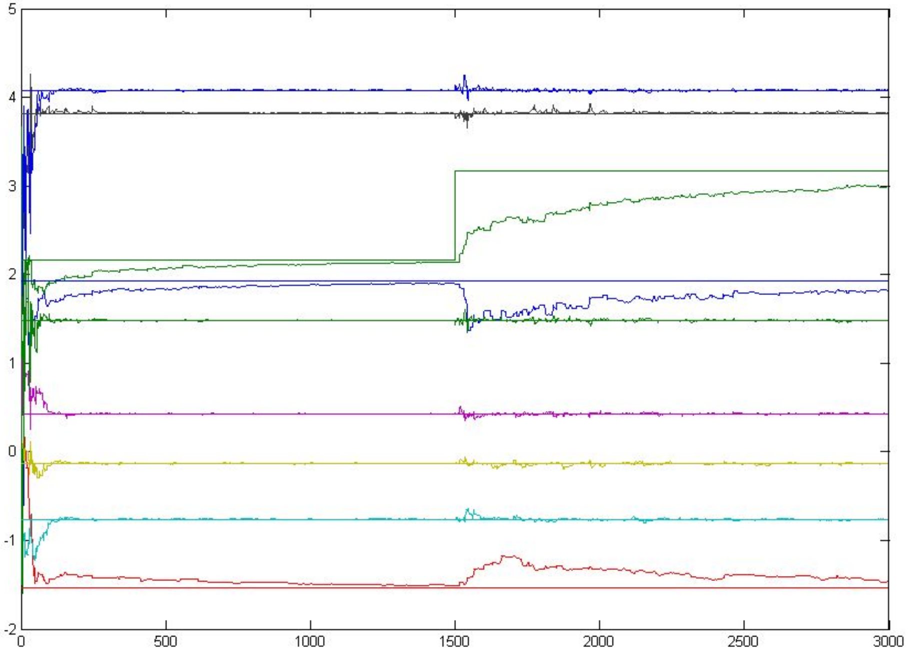
$W_{AD} = -0.3 + 0.00026 = -0.03;$

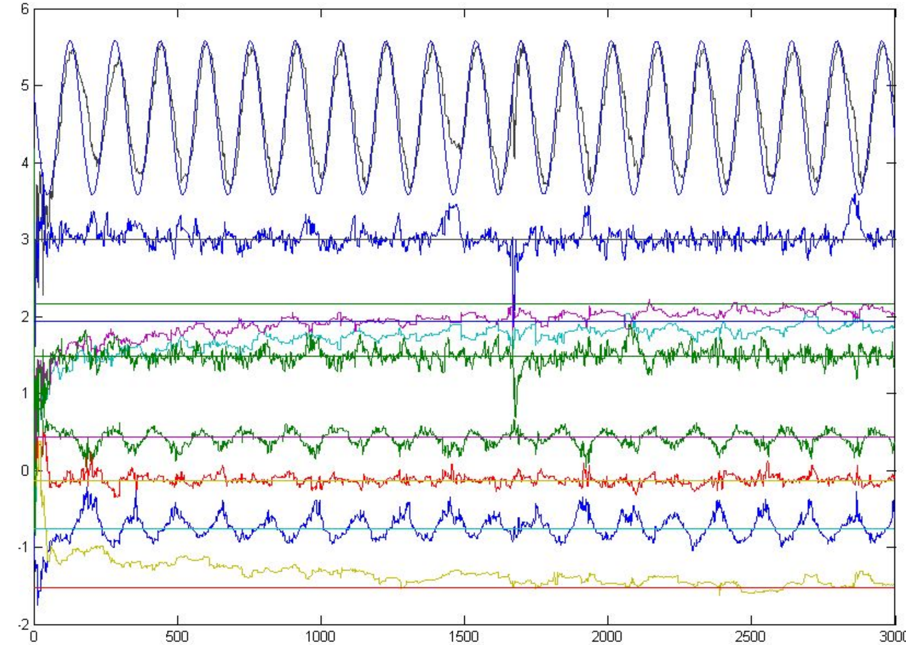
$W_{BC} = 0.1 + 0.00166 = 0.102;$

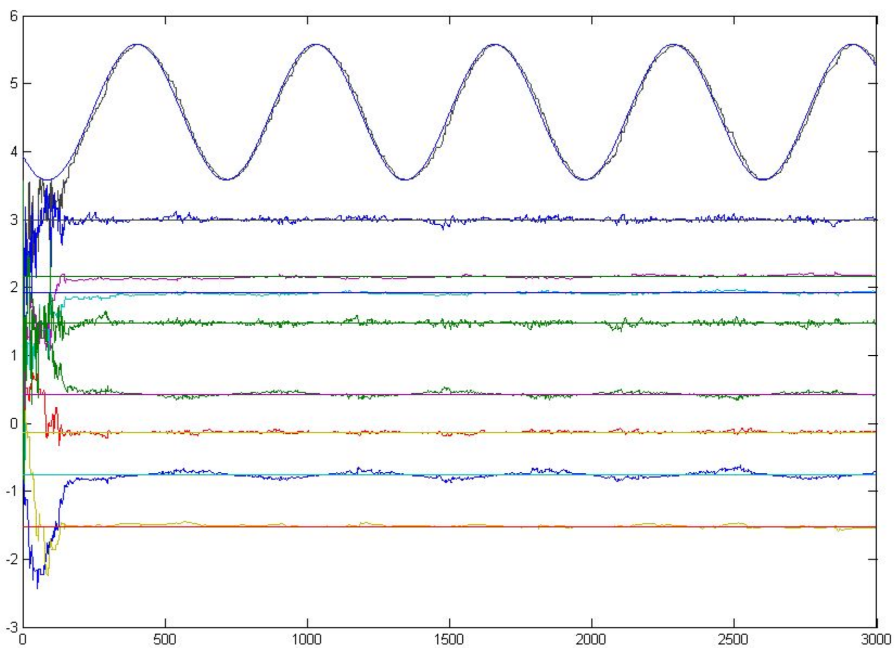
$W_{BD} = 0.3 - 0.00043 = -0.03;$

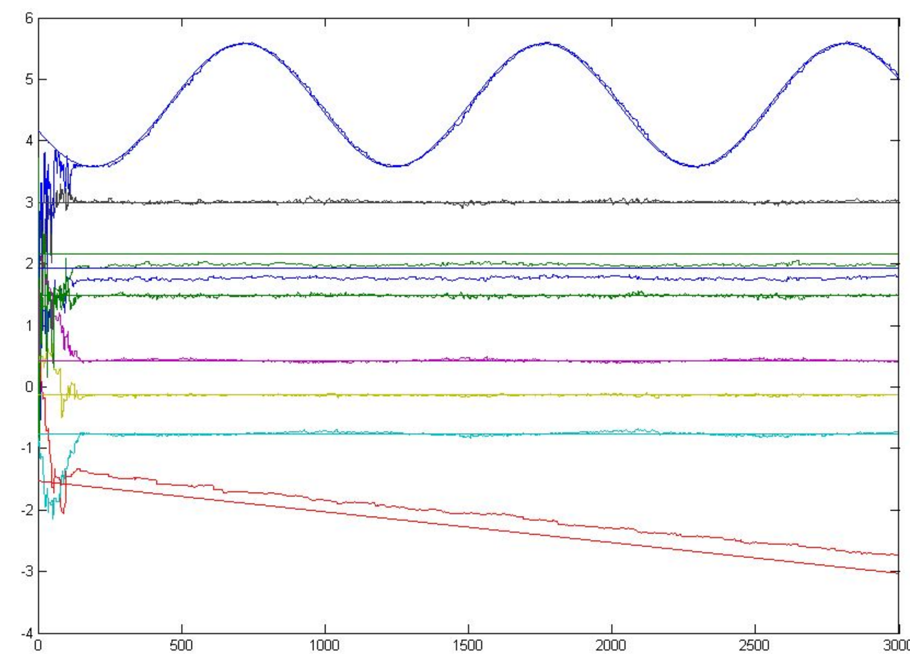
$W_{1C} = -0.4 + 0.00332 = -0.397;$

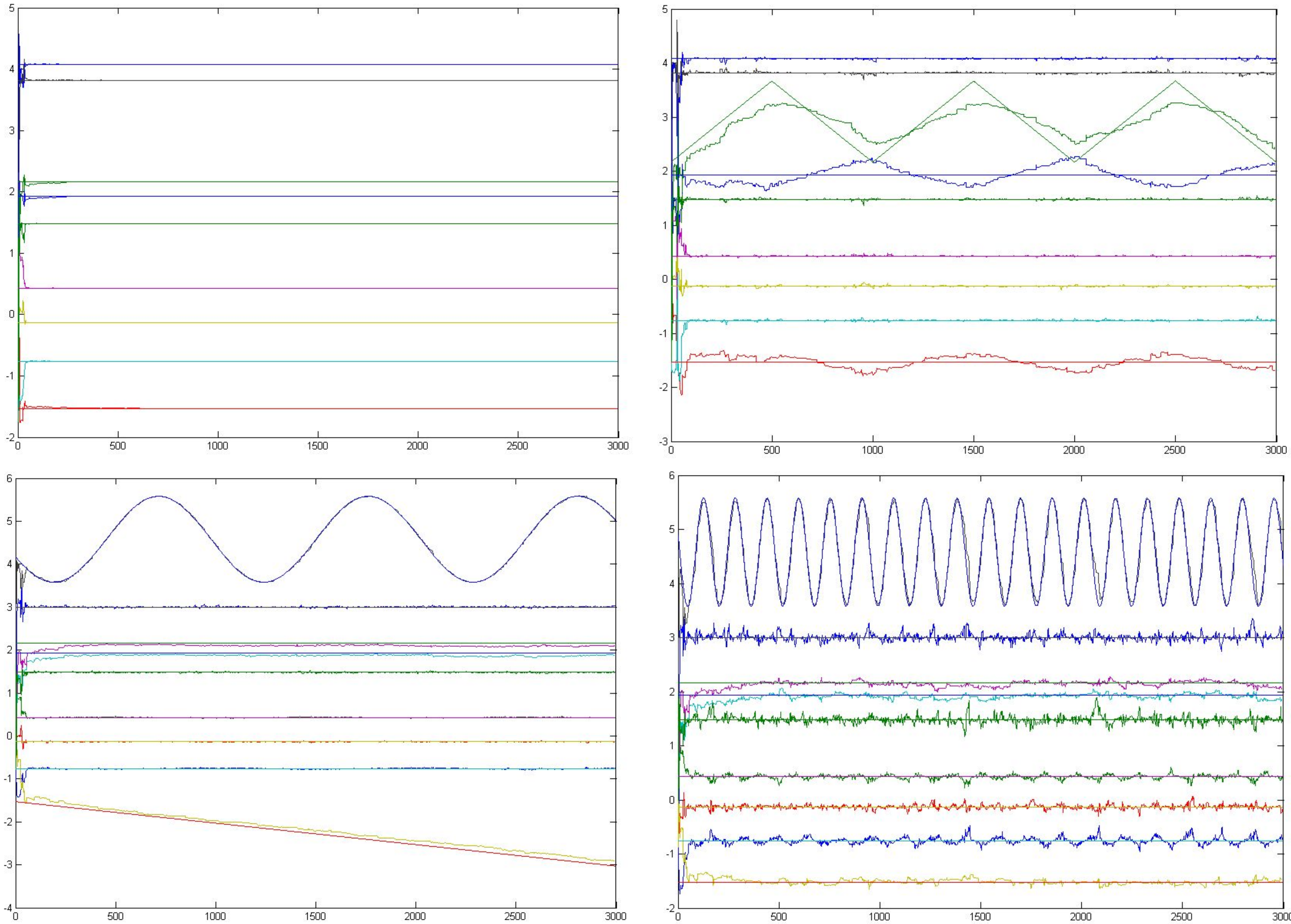
$W_{AD} = -0.3 - 0.00086 = -0.301.$

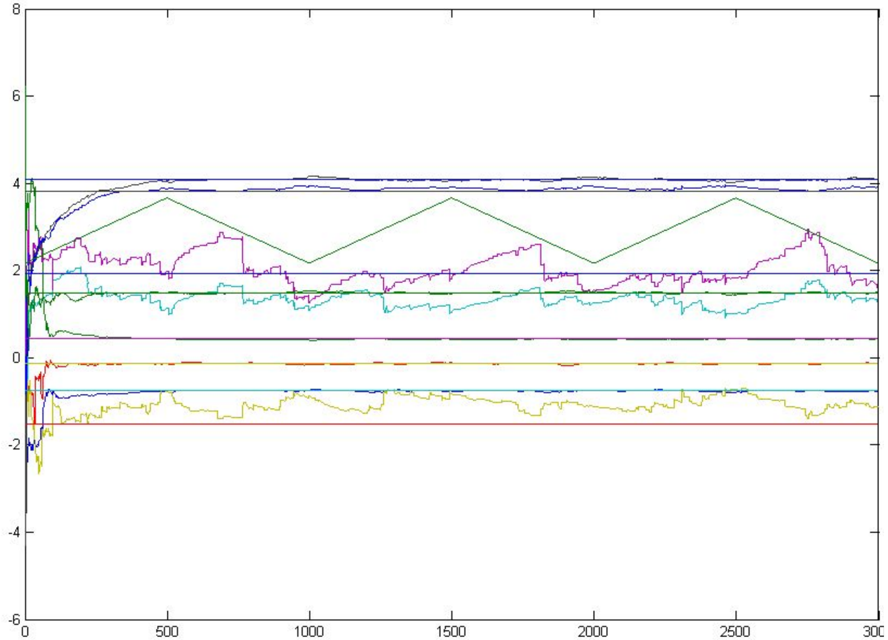

5
4
3
2
1
0
-1
-2
0 500 1000 1500 2000 2500 3000

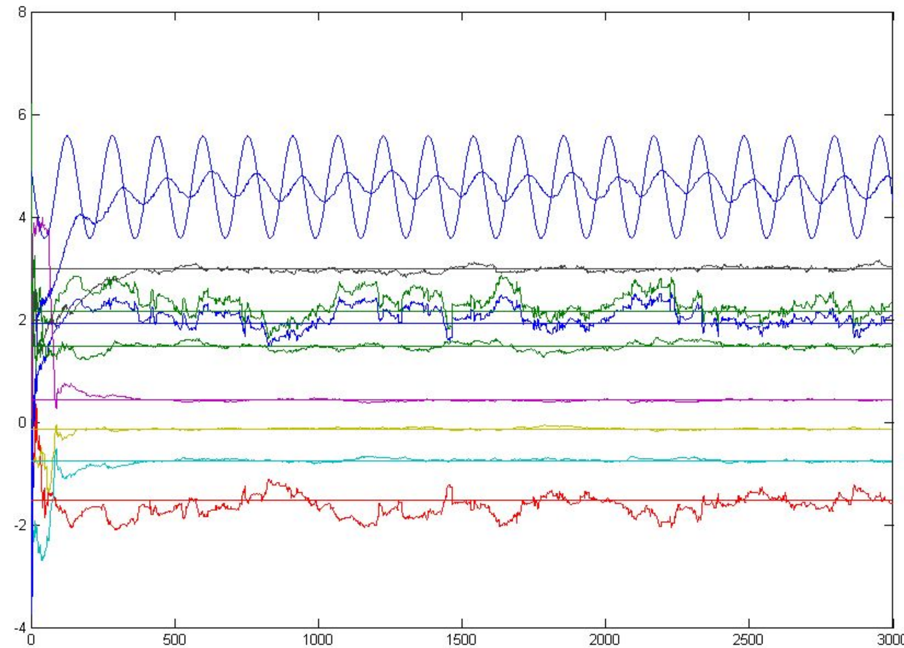

6
5
4
3
2
1
0
-1
-2
0 500 1000 1500 2000 2500 3000

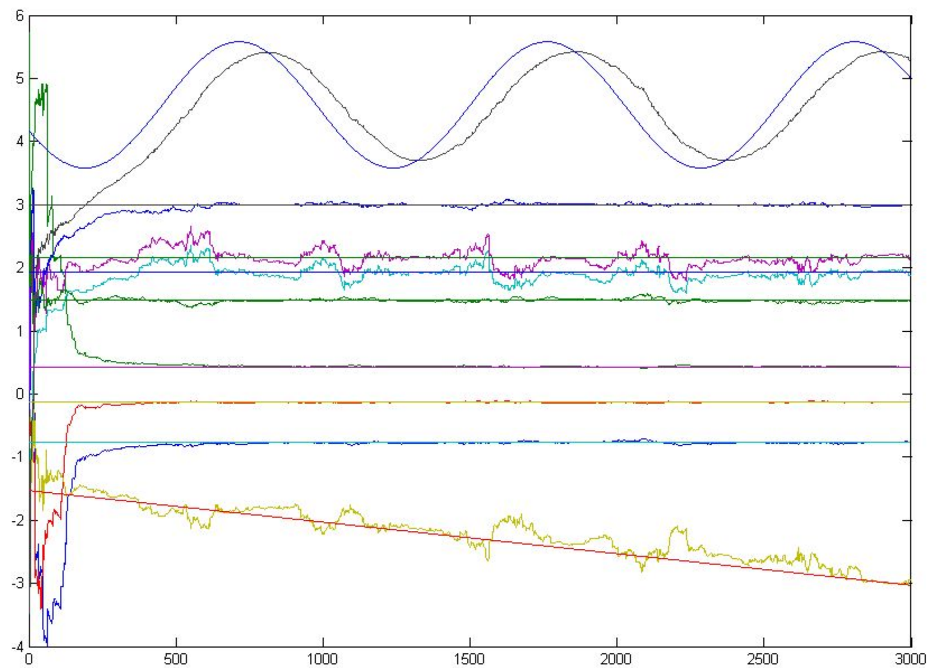

6
5
4
3
2
1
0
-1
-2
-3
0 500 1000 1500 2000 2500 3000


6
5
4
3
2
1
0
-1
-2
-3
-4
0 500 1000 1500 2000 2500 3000


5 4 3 2 1 0 -1 -2
0 500 1000 1500 2000 2500 3000
5 4 3 2 1 0 -1 -2 -3
0 500 1000 1500 2000 2500 3000
6 5 4 3 2 1 0 -1 -2 -3 -4
0 500 1000 1500 2000 2500 3000
6 5 4 3 2 1 0 -1 -2
0 500 1000 1500 2000 2500 3000


8
6
4
2
0
-2
-4
-6
0
500
1000
1500
2000
2500
300


8
6
4
2
0
-2
-4
0
500
1000
1500
2000
2500
3000


6
5
4
3
2
1
0
-1
-2
-3
-4
0
500
1000
1500
2000
2500
3000