TA 91.

Periodic signals in neural-like networks - an averaging analysis

R. Reinke[†] and D. Prätzel-Wolters[†]

[†] Arbeitsgruppe Technomathematik, Universität Kaiserslautern, Postfach 3049, 67653 Kaiserslautern, Germany
E-Mail: prwolt@mathematik.uni-kl.de, Tel.: +631-205-2338, Fax: +631-205-3052

Abstract - The paper describes the concepts and background theory of the analysis of a neural-like network for the learning and replication of periodic signals containing a finite number of distinct frequency components. The approach is based on a two stage process consisting of a learning phase when the network is driven by the required signal followed by a replication phase where the network operates in a autonomous feedback mode whilst continuing to generate the required signal to a desired accuray for a specified time.
The analysis focusses on stability properties of a model reference adaptive control based learning scheme via the averaging method. The averaging analysis provides fast adaptive algorithms with proven convergence properties.

Key Words - Periodic signals, neural-like networks, learning, replication, exponential stability, averaging

1. Introduction

The problem to be considered arises [6] in the development of models of learning of the repetitive motion of walking. However, in the paper a more general approach is taken to permit the application to the development of training algorithms for any periodic action such as met in control requirements for robotic manufacturing systems. In previous studies [10], emphasis has been placed on the use of systems theoretical concepts in the *development of algorithms with provable convergence properties*. Techniques included the use of appropriate canonical forms, Lyapunov stability theory and aspects of adaptive control convergence theory.
In this paper the proposed learning schemes (cf. [11]) are analyzed with the help of the averaging method in order to get estimates for the exponential convergence rates of the algorithms and improve the convergence properties by suitable selection of certain design parameters in the adaptive learning algorithms.

2. Problem Definition and learning rules

The precise problem to consider is the construction and training of a dynamic neural-like network whose aim is to "reconfigure" in order to replicate an arbitrary periodic signal $r(t)$ of a signal class R_N , whose elements are of the form:

$$r(t) = \sum_{k=1}^{N} A_k \sin(\omega_k t + \phi_k)\,,\ \omega_i \neq \omega_j \text{ for } i \neq j \tag{2.1}$$

$$1 \leq i,j \leq N,\ \omega_j \in \mathrm{IR}^+,\ \phi_j \in [0, 2\pi)\,,\ 0 < A_j \in \mathrm{IR} \tag{2.2}$$

The network is initially taken to be described by a state space model of the form

$$\frac{dx(t)}{dt} = A(\omega(t))x(t) + bu(t) \tag{2.3}$$

where $x(t) \in \mathrm{IR}^n$ and $u(t) \in \mathrm{IR}$ is an input to be defined. The network has the output

$$y(t) = c^T x(t) \tag{2.4}$$

where f^T denotes transpose of the vector f and $\omega(t) = (\omega_1(t), ..., \omega_{n_\omega}(t))^T \in \mathrm{IR}^{n_\omega}$ is a vector of time dependent "weights" to be adjusted continuously to achieve the certain objectives. The model can be regarded as a "linearization" of models of the form described in [3], for example, where the state dynamics have the typical structure

$$\frac{dx(t)}{dt} = \sigma(A(\omega(t))x(t) + bu(t)) \tag{2.5}$$

where $\sigma(\cdot)$ represents a vector-valued sigmoid function. The motivation for the use of the linear version is manifold and includes: the potential for the introduction of linear systems theory methods for performance and stability evaluation, the argument that an algorithm that does not work for the linear systems is unlikely to work for the nonlinear case and the periodic signals (2.1) are solutions of a linear differential equation. Although not conclusive, these motivations are sufficient to make study of the case described of interest.

The approach consists of two phases, one of learning and one of replication of the learned responses:

Phase One (THE LEARNING PHASE):

The input $u(t)$ is set equal to the desired output signal $r(t) \in R_N$ and the weights are adjusted in real time to ensure the success of the second phase. In effect, the objective is that during the system is excited by the required output, an (implicit) identification takes place and the output $y(t)$ asymptotically tracks the stimulus:

$$\lim_{t\to\infty} (r(t) - y(t)) = 0 \tag{2.6}$$

Phase Two (THE REPLICATION PHASE):

After a suitable period of time T^*, $u(t)$ is switched/replaced by the network output $y(t)$, the weight vector $\omega(t)$ is frozen at its value at $t = T^*$ and the resultant time invariant (positive feedback closed loop) system tracks the (decoupled) stimulus to the desired acccuracy for the desired period T_r, i.e. in the time interval $[T^*, T^* + T_r]$. More precisely the requirement for the replication phase is:
For every time period $T_r > 0$ and replication accuracy $\varepsilon > 0$ there exists a switching time $T^*(\varepsilon, T_r) < \infty$ sucht that the reponse of the frozen system

$$x(t) = (A(\omega(T^*)) + bc^T)x(t), \; y(t) = c^T x(t) \tag{2.7}$$

satisfies:

$$|y(t) - r(t)| < \varepsilon \text{ for all } t \in [T^*, T^* + T_r].$$

(2.6) is required for all $r(\cdot) \in R_N$ and the adjustment of the weights have to be formed under the restricted knowledge of the particular signal $r(\cdot)$ itself, the frequences ω_j, amplitudes A_j and phase shifts ϕ_j are unknown, hence not available for the adaptation procedure.
In [10] is is shown that this objectives can be achieved with a linear network of the form:

network:
$$\begin{aligned} \dot{x}(t) &= A(\omega(t))x(t) + bu(t) \\ y &= c^T x(t) \end{aligned}$$

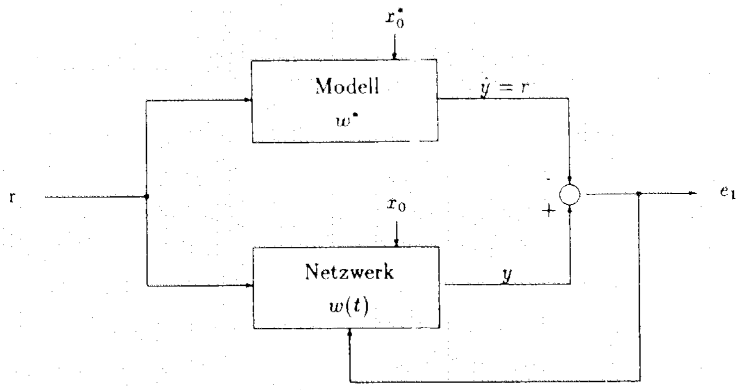
where

$$A(\omega) = \begin{bmatrix} -\alpha & 1 & 0 & 0 & \cdots & \cdots & 0 \\ 0 & -\alpha & 1 & 0 & \cdots & \cdots & 0 \\ \vdots & & \ddots & \ddots & & & \vdots \\ \vdots & & & \ddots & \ddots & & \vdots \\ \vdots & & & & \ddots & 1 & 0 \\ 0 & & & & & -\alpha & 0 \\ \omega_1 & \cdots & \cdots & \cdots & \cdots & \omega_{n-1} & -\alpha \end{bmatrix}, n \times n \qquad \begin{aligned} \alpha &> 0 \\ b &= [0, ..., 0, 1, 0]^T \\ c &= [0, ..., 0, 1]^T \end{aligned} \tag{2.8}$$

and state dimension $n = 2N + 1$.
The learning scheme is based on a model reference adaptive control approach (cf. Fig. 1), where for every $r(\cdot) \in R_N$ the model is given by the equations:

model:
$$\begin{aligned} \dot{\hat{x}}(t) &= A(\omega^*)\hat{x}(t) + bu(t) \quad , \; \hat{x}(0) = x_o^* \\ \hat{y} &= c\hat{x}(t) \end{aligned} \tag{2.9}$$

where the fixed ideal parameters ω^* and x_o^* are chosen s.t. the output of (2.9) becomes identifical to its input for $u(t) = r(t)$ and for all $t \geq 0$.



MRAC-approach

Defining the symbols $e(t) = x(t) - \hat{x}(t)$, the state error, $e_P(t) = \omega(t) - \omega^*$, the parameter error, then leads to the error dynamics

$$\dot{e}(t) = A(\omega^*)\,e(t) + cx^T(t)Re_p(t) \tag{2.10}$$

where

$$R^T = [I_{2N}, 0_{2N,1}]\,.$$

The approach to the problem uses the Liapunov function candidate

$$V(e, e_P) = e^T P_e + e_P^T Q^{-1} e_P \tag{2.11}$$

where $0 < Q = Q^T \in \mathbb{R}^{2N\times 2N}$ is arbitrary and $P = P^T > 0$ is the unique solution of the Liapunov equation

$$A^T(\omega^*)P + PA(\omega^*) = -S \tag{2.12}$$

with $S \in \mathbb{R}^{(2N+1)\times(2N+1)}$ arbitrarily, symmetric positive definite. With the adaption /learning law:

$$\dot{w}(t) = -QR^T x(t) c^T Pe \tag{2.13}$$

the following result is proved [cf. [11]] :

Theorem 2.1:

Let $Q \in \mathbb{R}^{2N\times 2N}$ and $S \in \mathbb{R}^{(2N+1)\times(2N+1)}$ be positive definite symmetric matrices and $P \in \mathbb{R}^{(2N+1)\times(2N+1)}$ be the unique solution of Liapunov equation (2.12). Then the combined error system of (2.10) and (2.13)

$$\begin{bmatrix} e(t) \\ e_p(t) \end{bmatrix}^{\bullet} = \begin{bmatrix} A(\omega^*) & cx^T(t)R^T \\ -QRx(t)c^TP & 0 \end{bmatrix} \begin{bmatrix} e(t) \\ e_p(t) \end{bmatrix} \tag{2.14}$$

is exponentially stable, in particular for all initial conditions $x(0)$ and $\omega(0)$ we have:

$$\lim_{t\to\infty} e(t) = 0 \quad \text{and} \quad \lim_{t\to\infty} \omega(t) = \omega^*$$

with exponential convergence. □

Simulations show (c.f. section 4) that the convergence speed, hence the exponential convergence rate of (2.14) heavily depends on the selection of the matrix Q. The purpose of the following two sections is an estimation of this convergence rate and the development of a procedure for suitable selection of Q which results in fast learning algorithms.

3. Averaging

Averaging is a method for the analysis of differential equations of the form

$$\dot{x} = \varepsilon f(t,x)$$

with the aim, to relate the solutions of (3.1) for ε suffiently small with the solutions of the averaged system

$$\dot{x}_{av} = \varepsilon f_{av}(a_{av}) \text{ , where } f_{av}(x) = \lim_{T\to\infty} \frac{1}{T} \int_{t_o}^{t_o+T} f(\tau,x)d\tau$$

This method was introduced by Bogoliuboff and Mitropolskii [4] and developed further by Volosov [14], Sethna [13], Balachandra and Sethna [2], Hale [9] and Sanders and Verhulst [12]. More geometrically oriented formulations can be found in Arnold [1] and Guckenheimer and Holmes [8].
Here we consider systems of differential equations of the form

$$\dot{x} = \varepsilon f(t,x,y) \text{ , } x(0) = x_o \tag{3.1a}$$

$$\dot{y} = Ay + \varepsilon g(t,x,y) \text{ , } y(0) = y_o \tag{3.1b}$$

where $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$, $0 < \varepsilon < \varepsilon_o$ and $f(t,0,0) = 0$ and $g(t,0,0) = 0$ for all $t \geq 0$, i.e. $(x,y) = (0,0)$ is an equilibrium point of (3.1). Furthermore we assume that the matrix of the fast dynamics (3.1b) is asymptotically stable.

The aim of the averaging method is to relate the slow dynamics (3.1a) with the dynamics of an averaged system

$$\dot{x}_{av} = \varepsilon f_{av}(x_{av}), \; x_{av}(0) = x_o. \tag{3.2}$$

More precisely as above the averaged function $f_{av}(x)$ is defined by:

$$f_{av}(x) = \lim_{T\to\infty} \frac{1}{T} \int_{t_o}^{t_o+T} f(\tau, x, 0)\, d\tau \tag{3.3}$$

Definition 3.1:
A function $f(t,x,0)$ has the average $f_{av}(x)$ iff there exists a function $\gamma(T) : \mathbb{R}^+ \to \mathbb{R}^+$, continuous and strong monotonically decreasing with $\lim_{T\to\infty} \gamma(T) = 0$ such that

$$\left| \frac{1}{T} \int_{t_o}^{t_o+T} f(\tau, x, 0)\, d\tau - f_{av}(x) \right| \le \gamma(T) \tag{3.4}$$

for all $t_o \ge 0, T > 0$ and $x \in \{x \in \mathbb{R}^n, \; |x| \le h\} =: B_h$

□

$\gamma(T)$ is called convergence function.

In Fu et al [7] the following theorem is proved:

Theorem 3.2:
The system (3.1) has an exponentially stable equilibrium point 0 for all ε sufficiently small if the following conditions are satisfied:

(i) f and g are piecewise continuous in t and there exist constants $\ell_1, \ell_2, \ell_3, \ell_4$ and $\ell_{av} \ge 0$ such that

$$|f(t,x_1,y_1) - f(t,x_2,y_2)| \le \ell_1 |x_1 - x_2| + \ell_2 |y_1 - y_2| \tag{3.5}$$

$$|g(t,x_1,y_1) - g(t,x_2,y_2)| \le \ell_3 |x_1 - x_2| + \ell_4 |y_1 - y_2| \tag{3.6}$$

$$|f_{av}(x_1) - f_{av}(x_2)| \le \ell_{av} |x_1 - x_2| \tag{3.7}$$

for all $x_1, x_2 \in B_h, \; y_1, y_2 \in B_h$ and $t \ge 0$.

(ii) The function $d(t,x) := f(t,x,0) - f_{av}(x)$ is continuously differentiable with respect to x, has mean value 0 with convergence function $\gamma(T) \cdot |x|$ and $\frac{\partial d(t,x)}{\partial x}$ is bounded with mean value 0 and convergence function $\gamma(T)$.

□

System (3.1) has two seperate time scales in the sense that (3.1b) is independent of (3.1a) if $\varepsilon \to 0$. However the dynamics (2.14) considered in this paper have two mixed time scales. They can be considered as a special case of:

$$\dot{x} = \varepsilon \tilde{f}\,(t, x, \tilde{y})\,,\ x(0) = x_o \tag{3.8a}$$

$$\dot{\tilde{y}} = A\tilde{y} + h(t, x) + \varepsilon\tilde{g}(t, x, \tilde{y}),\ \tilde{y}(0) = y_o \tag{3.8b}$$

In order to apply theorem 3.2 this system has to be transformed into the form (3.1). For this $h(t, x)$ is transformed into

$$\ell(t, x) := \int_0^t e^{A(t-\tau)} h(\tau, x) d\tau \tag{3.9}$$

and then in (3.3) the modified mean value function

$$f_{av}(x) = \lim_{T\to\infty} \frac{1}{T} \int_{t_o}^{t_o+T} \tilde{f}(\tau, x, \ell(\tau, x)) d\tau \tag{3.10}$$

is selected. Via the L-transformation

$$y := \tilde{y} - \ell\,(t, x)$$

(3.8) admits the form (3.1) with:

$$f(t, x, y) := \tilde{f}(t, x, y + \ell(t, x)) \tag{3.11}$$

and

$$g(t, x, y) := -\frac{\partial \ell(t,x)}{\partial x} \tilde{f}(t, x, y + \ell(t, x)) + \tilde{g}(t, x, y + \ell(t, x)) \tag{3.12}$$

3.1 Analysis of the error system (2.13) by the averaging-method

In order to apply the averaging method to our error system

$$\dot{e}_p(t) = -\varepsilon Q v(t) c^T P e(t) \tag{3.13}$$

$$\dot{e}(t) = A(\omega^*) e(t) + c v^T(t) e_P(t) \tag{3.14}$$

with $v(t) = R^T x(t)$ we identify $x \mathrel{\hat{=}} e_P$ and $\tilde{y} \mathrel{\hat{=}} e$ and obtain:

$$f(t, x, y) = -Q v(t) c^T P(y + L(t)x)$$

$$g(t, x, y) = L(t) Q v(t) c^T P(y + L(t)x)$$

where

$$\ell(t, x) = L(t)x := \int_0^t e^{A(\omega^*)(t-\tau)} c v^T(\tau) d\tau \cdot x.$$

The averaged system of (3.13) and (3.14) is:

$$\dot{x}_{av} = \varepsilon f_{av}(x_{av})$$

with

$$f_{av}(x) = -\lim_{T\to\infty} \frac{1}{T} \int_{t_o}^{t_o+T} Qv(t)c^T PL(t)x dt \tag{3.15}$$

$$= \lim_{T\to\infty} \frac{1}{T} \int_{t_o}^{t_o+T} Qv(t)c^T P \int_{t_o}^{t} e^{A(\omega^*)(t-\tau)} cv^T(\tau) d\tau dt. \tag{3.16}$$

Obviously we have $f(t,0,0) = 0$ and $g(t,0,0) = 0$, hence $(e, e_P) = 0$ is an equilibrium point of (3.13) and (3.14) and $A(\omega^*)$ is asymptotically stable. Furthermore:

Lemma 3.3:

The error system (3.13) and (3.14) satisfies the conditions (i) and (ii) of theorem 3.2 for all $h > 0$.

Proof:

$v(t)$ and $L(t)$ are bounded and the partial derivatives in

$$\begin{pmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \\ \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \end{pmatrix} = \begin{pmatrix} -Qv(t)c^T PL(t) & -Qv(t)c^T P \\ L(t)Qv(t)c^T PL(t) & L(t)Qv(t)c^T p \end{pmatrix}$$

are continuous and bounded, which implies the first two Lipschitz conditions in (i). Furthermore $f_{av}(0) = 0$ and

$$\frac{\partial f_{av}}{\partial x_{av}} = -\lim_{T\to\infty} \frac{1}{T} \int_{t_o}^{t_o+T} Qv(t)c^T P \int_{t_o}^{t} e^{A(\omega^*)(t-\tau)} cv^T(\tau) d\tau dt$$

is continuous and bounded, hence the third Lipschitz condition in (i) is satisfied and

$$\frac{\partial}{\partial x} d(t,x) = \frac{\partial f(t,x,0)}{\partial x} - \frac{\partial f_{av}(x)}{\partial x}$$

is continuous and bounded too. The convergence function $\gamma(T)$ of $\frac{\partial d(t,x)}{\partial x}$ is the convergence function of $\frac{\partial f(t,x,0)}{\partial x}$. $v(t)$ and $L(T)$ are generated by filtering of the periodic signal $r(t)$ with an asymptotically stable linear system, hence $\frac{\partial f(t,x,0)}{\partial x}$ is up to an asymptotically vanishing term identical to a finite sum of periodic function. By a lemma in Sanders and Verhulst [12] this implies that the convergence function $\gamma(T)$ is of order $O\left(\frac{1}{T}\right)$. From $f(t,x,0) = \frac{\partial}{\partial x} f(t,0,x)x$ we conclude that the convergence function of $f(t,x,0)$ is of the form $\gamma(T) \cdot |x|$ □

From (3.1b) we obtain for the averaged system the representation:

$$\dot{e}_{p_{av}} = -\varepsilon \cdot A_{av} \cdot e_{p_{av}},\ e_{p_{av}}(0) = e_p(0) \tag{3.17}$$

with

$$A_{av} = Q \cdot \lim_{T\to\infty} \frac{1}{T} \int_{t_o}^{t_o+T} v(t)c^T P \int_{t_o}^{t} e^{A(\omega^*)(t-\tau)} c v^T(\tau) d\tau dt. \tag{3.18}$$

However, Q is positive definite and therefore the averaged system (3.17) is exponentially stable iff the eigenvalues of the linear operator

$$\lim_{T\to\infty} \frac{1}{T} \int_{t_o}^{t_o+T} v(t) \cdot \tilde{v}^T(t) d\tau$$

where

$$\tilde{v}^T(t) := c^T P \int_{t_o}^{t} e^{A(\omega^*)(t-\tau)} c v^T(\tau) d\tau$$

belong to $\mathbb{C}^+$.

For the analysis of this operator we need some concepts of Generalized Harmonic Analysis.

Definition 3.4:

(i) A function $u : \mathbb{R}^+ \to \mathbb{R}^n$ is called stationary if the limit

$$Cov_u(\tau) := \lim_{T\to\infty} \frac{1}{T} \int_{t}^{t+T} u(s) u^T(\tau + s) ds$$

exists uniformly with respect to t. $Cov_u(\tau)$ is called autocovariance of $u(\cdot)$.

(ii) If $Cov_u(\cdot)$ is continuous then

$$Cov_u(\tau) = \int_{-\infty}^{\infty} e^{iw\tau} S_u(dw)$$

where $S_u(dw)$ is called the spectral measure of $u(\cdot)$. The integral $(\cdot)$ exists by Bochners theorem (cf. Bochner [3]), because $Cov_u(\cdot)$ is a positive semidefinite function.

(iii) If $u : \mathbb{R}^+ \to \mathbb{R}^m$ and $y : \mathbb{R}^+ \to \mathbb{R}^p$ are stationary, then the limit

$$Cov_{y,u}(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_{t}^{t+T} y(s) u^T(\tau + s) ds$$

exists uniformly in T and is called the covariance of u and y. □

The relation between the auto- and covariances of the input and output signals of a stable time-invariant linear system is described by the linear filter lemma of Boyd and Sastry [5] :

Lemma 3.5: (Linear filter lemma)

Let $H(s)$ be a stable, strictly proper rational transformmatrix with impuls response $h(t) \in \mathbb{R}^{p\times m}$, $\hat{y}(s) = H(s)\hat{u}(s)$ and let $u(\cdot)$ be stationary. Then

(i) y is stationary with autocovariance

$$Cov_y(t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(\tau_1)Cov_u(t+\tau_1-\tau_2)\cdot h^T(\tau_2)\,d\tau_1 d\tau_2$$

and spectral measure

$$S_y(dw) = H(iw)S_u(dw)H^*(iw).$$

(ii)

$$Cov_{y,u}(t) = \int_{-\infty}^{\infty} h(\tau_1)Cov_u(t+\tau_1)d\tau_1$$

$$S_{yu}(dw) = H(iw)S_u(dw).$$ □

Based on the above notations the dynamic matrix A_{av} of the averaged system can be expressed as:

$$A_{av} = Q \cdot Cov_{v,\tilde{v}}(0).$$

Lemma 3.6:

$$\sigma\left(Cov_{v,\tilde{v}}(0)\right) \in \mathbb{C}^+ \tag{3.19}$$

Proof:

(3.18) leads to:

$$\hat{\tilde{v}}(s) = M(s)\cdot\hat{v}(s)$$

with

$$M(s) = c^T P(sI - A(\omega))^{-1}c = \tfrac{p_{n,n}}{s+\alpha}$$

and

$$p_{n,n} > 0.$$

However

$$\begin{aligned}\hat{v}(s) &= R^T\hat{x}(s)\\ &= R^T(sI - A(\omega^*))^{-1}b\cdot r(s)\\ &= H(s)\cdot r(s)\end{aligned}$$

where

$$H(s) = \left[\tfrac{1}{(s+\alpha)^{n-1}}, \tfrac{1}{(s+\alpha)^{n-2}}, \dots \tfrac{1}{s+\alpha}\right]^T$$

The linear filter lemma implies:

$$Cov_{v,\tilde{v}}(0) = \int_{-\infty}^{\infty} H(i\omega)S_r(d\omega)H^*(i\omega)M^*(i\omega).$$

The covariance of $r(\cdot)$ is:

$$Cov_r(\tau) = \sum_{j=1}^{k}\frac{A_j^2}{2}\cos(\omega_j\tau)$$

with associated spectral measure:

$$S_r(d\omega) = \sum_{j=1}^{k} \frac{A_j^2}{2}(\delta_{-\omega_j} + \delta_{\omega_j}).$$

Hence:

$$Cov_{v,\tilde{v}}(0) = \sum_{j=1}^{k} \frac{A_j^2}{4}(H(\text{-}i\omega_j)H^*(\text{-}i\omega_j)M^*(\text{-}i\omega_j) + H(i\omega_j)H^*(i\omega_j)M^*(i\omega_j)) \in \mathbb{R}^{n-1\times n-1} \tag{3.20}$$

Every summand in (3.20) satisfies:

(i)
$$\begin{aligned} F(i\omega_j) &:= H(-i\omega_j)H^*(-i\omega_j)M^*(-i\omega_j) + H(i\omega_j)H^*(i\omega_j)M^*(i\omega_j) \\ &= 2Re\{H(i\omega_j)H^*(i\omega_j)\}\, Re\{M(i\omega_j)\} + 2Im\{H(i\omega_j)H^*(i\omega_j)\}\, Im\{M(i\omega_j)\} \end{aligned}$$

(ii) $rk F(i\omega_j) = 2$,

(iii)
$$\begin{aligned} F(i\omega_j) + F^T(i\omega_j) &= H(-i\omega_j)H^*(-i\omega_j)M(i\omega_j) + H(i\omega_j)H^*(i\omega_j)M(-i\omega_j) \\ &\quad + H(i\omega_j)H^*(i\omega_j)M(i\omega_j) + H(-i\omega_j)H^*(-i\omega_j)M(-i\omega_j) \\ &= 4Re\{H(i\omega_j)H^*(i\omega_j)\}\, Re\{M(i\omega_j)\}. \end{aligned}$$

$F(i\omega_j) + F^T(i\omega_j)$ is positive semidefinite because $Re\{M(i\omega)\} > 0$. $H(i\omega_j)$ and $H(i\omega_j)$ are linear independent for $\omega_j \neq \omega_\ell$, $j \neq \ell$. This together with (ii) and (iii) implies:

$$rk\ Cov_{v,\tilde{v}}(0) = \min(n-1, 2k)$$

hence $Cov_{v,\tilde{v}}(0) + Cov_{v,\tilde{v}}^T(0)$ is positive definite for $n = 2k + 1$and

$$\sigma(Cov_{v,\tilde{v}}(0)) \subset \mathbb{C}^+. \qquad \square$$

Now note that by theorem 3.2 the original system (3.13) and (3.14) is exponentially stable for $\varepsilon > 0$ sufficiently small, i.e. for a sufficiently slow adaption rate.

4. Approximation of exponential convergence rate and selection of design parameter Q

From the proof of theorem 3.2 cf. Fu et al.([7]) it is easy to verify that if the convergence function $\gamma(T)$ is of $0\left(T^{-r}\right)$, $0 < r \leq 1$ and the exponential convergence rate of the averaged system is $\varepsilon \cdot \alpha_{av}$ then the convergence rate of the original system is at least $\varepsilon \cdot \alpha_{av}\left(1 - 0\left(\varepsilon^{1/3} + \varepsilon^r\right)\right)$. In particular for $\varepsilon \to 0$ this convergence rate becomes identical to $\varepsilon \cdot \alpha_{av}$, the convergence rate of the averaged system.

This implies that for $\varepsilon > 0$ small enough the convergence rate γ of our error systems is approximated by

$$\gamma_{av} = \varepsilon \cdot Re\left(\lambda_{\min}(QCov_{v,\tilde{v}}(0)\right),$$

i.e. there exists a constant $m > 0$, s.t.

$$|e(t), e_p(t)| \leq m e^{-\gamma_{av}(t-t_o)}.$$

Given this result the following strategy for the selection of the positive definite design matrix Q leads to a fast convergence speed of the learning algorithm: Select Q in such a way, that the reals parts of $QCov_{v,\tilde{v}}(0)$ are approximately all the same.

For the selection $Q = I$ the real parts of the eigenvalues of $Cov_{v,\tilde{v}}(0)$ determine the convergence rate. However the sizes of these eigenvalues differ strongly for large netdimensions. A better selection is

$$Q = Cov_v(0)^{-1}$$

Consider for example the simplest net dimensions $n = 3$ and $n = 5$:

(i) $k = 1,\ n - 1 = 2,\ r = \sin(\omega t + \varphi)$. Then

$$\begin{aligned} T &= Cov_v(0)^{-1} Cov_{v,\tilde{v}}(0) \\ &= \begin{bmatrix} \frac{2\alpha}{\omega^2+\alpha^2} & 1 \\ -\frac{1}{\omega^2+\alpha^2} & 0 \end{bmatrix} \end{aligned}$$

$$\sigma(T) = \left\{ \frac{\alpha}{\omega^2+\alpha^2} \pm i \frac{\omega}{\omega^2+\alpha^2} \right\}$$

(ii) $k = 2,\ n - 1 = 4,\ r = \sin(\omega_1 t + \varphi_1) + \sin(\omega_2 t + \varphi_2)$. Then

$$T = \begin{bmatrix} \frac{2\alpha(2\alpha^2+\omega_1^2+\omega_2^2)}{(\omega_1^2+\alpha^2)(\omega_2^2+\alpha^2)} & 1 & 0 & 0 \\ -\frac{6\alpha^2+\omega_1^2+\omega_2^2}{(\omega_1^2+\alpha^2)(\omega_2^2+\alpha^2)} & 0 & 1 & 0 \\ \frac{4\alpha}{(\omega_1^2+\alpha^2)(\omega_2^2+\alpha^2)} & 0 & 0 & 1 \\ -\frac{1}{(\omega_1^2+\alpha^2)(\omega_2^2+\alpha^2)} & 0 & 0 & 0 \end{bmatrix}$$

$$\sigma(T) = \left\{ \frac{\alpha}{\omega_j^2+\alpha^2} \pm i \frac{\omega_j}{\omega_j^2+\alpha^2}\ ,\ j = 1, 2 \right\}.$$

Without proof we

<u>Claim 4.1:</u>

(i) $\sigma\left(Cov_v(0)^{-1} Cov_{v,\tilde{v}}(0)\right) = \left\{ Re M(\pm i\omega_j) + i \cdot A_j^{-2} Im M(\pm i\omega_j),\ 1 \leq j \leq k \right\}$ where

$$M(s) = \frac{p_{n,n}}{s+\alpha}.$$

(ii) For $p_{n,n} = 1$

$$\gamma_{av} = \varepsilon \cdot \frac{\alpha}{\omega_{\max}^2 + \alpha^2}, \quad \omega_{\max} := \max_{1 \leq j \leq k} \omega_j$$

The effect of the selection of $Q = Cov_v(0)^{-1}$ instead of $Q = I$ is demonstrated by the following example.

Example 4.2:

Let $k = 2$, $n = 5$, $\alpha = 1$ and $r = \sin t + \sin 2t$. The eigenvalues of $Cov_{v,\tilde{v}}(0)$ are $\{0.1287 \pm 0.1037i,\ 0.0015,\ 0.0005\}$, however the eigenvalues of $Cov_v(0)^{-1} Cov_{v,\tilde{v}}(0)$ are $\{0.5 \pm 0.5i,\ 0.2 \pm 0.4i\}$, which leads to an approximation of the convergence rate $\gamma \approx 0,0005\varepsilon$ in the first case and $\gamma \approx 0,2\varepsilon$ in the second case. Hence in this case a substantial convergence acceleration can be expected. This is demonstrated by the simulations in section 5.

An apriori selection of Q as $Cov_v(0)^{-1}$ requires knowledge of the amplitudes and frequencies of the signal $r(\cdot)$, which are assumed to be not available for the network design. However $Cov_v(0)^{-1}$ can be computed adaptively from the available signal $v = R^T x$. This is possible because $Cov_v(0)$ satisfies the matrix differential equation

$$\dot{Q}(t) = -\tfrac{1}{t} Q(t) + \tfrac{1}{t} R^T x(t) x^T(t) R$$

which is in this form not suitable for the computation of $Cov_v(0)$ because of the pole at $t = 0$ of the right hand side.

However the modification

$$\dot{Q}(t) = \begin{cases} 0\,,\ Q(0) = I & ,\quad 0 \leq t < t_s \\ -\tfrac{1}{t} Q(t) + \tfrac{1}{t} R^T x(t) x(t)^T R & ,\quad t \geq t_s \end{cases} \tag{4.1}$$

has the solution

$$Q(t) = \frac{t_s I - \int_0^t R^T x(\tau) x^T(\tau) R d\tau + \int_0^t R^T(\tau) x^T(\tau) R d\tau}{t}$$

with

$$Q(t) \overset{t \to \infty}{\longrightarrow} Cov_{R^T x}(0).$$

The error $|Q(t) - Cov_{R^T x}(0)|$ is of order $0\left(\frac{1}{t}\right)$ and $Q(t)$ is positive definite for all $t > 0$.

This leads to the accelerated adaption law:

$$\dot{\omega} = -Q^{-1}(t) R^T x(t) (y(t) - r(t))$$

with $Q(t)$ from (4.1).

Application of the averaging method then leads to the following result:

Theorem 4.2

Given the net structure (2.8) with input $u = r = \sum_{j=1}^{N} A_j \sin(\omega_j t + \phi_j) \in R_N$, $\varepsilon > 0$ sufficiently small, then for the accelerated Liapunov learning rule:

$$\dot{\omega} = -\varepsilon Q(t)^{-1} R^T x(t)(y(t) - r(t)),\ \omega(0) = \omega_o \tag{4.2}$$

$$\dot{Q}(t) = \begin{cases} 0\ ,\ Q(0) = I & ,\ 0 \le t < t_s \\ -\frac{1}{t}Q(t) + \frac{1}{t}R^T x(t) x^T(t)^T R & ;\ t \ge t_s \end{cases} \tag{4.3}$$

the conclusions of theorem 2.1 remain true.

Proof:

Analogous to (3.13) and (3.14) the averaged system becomes

$$\dot{e}_{p_{av}} = -\varepsilon A_{av} e_{p_{av}}\ ,\ e_{p_{av}}(0) = e_p(0)$$

with

$$A_{av} = \lim_{T\to\infty} \frac{1}{T} \int_{t_o}^{t_o+T} Q^{-1}(t) v(t) c^T P \int_{t_o}^{t} e^{A(\omega^*)(t-\tau)} c v^T(\tau) d\tau dt$$

and

$$\lim_{t\to\infty} Q(t) = Cov_v(0)$$

implies

$$A_{av} = Cov_v(0)^{-1} Cov_{v,\tilde{v}}(0).$$

□

Remark 4.3:

An alternative adaption law with a nonsingular right hand side is

$$\dot{Q}(t) = -\rho Q(t) + \rho R^T x(t) x^T(t) R,\ Q(0) = I,\ \rho > 0. \tag{4.4}$$

(4.4) can be interpreted as a filtering every component signal of $R^T x(t) x^T(t) R$ with the stable linear filter $\frac{\rho}{s+\rho}$. The stationary state of (4.4) is an oscillation around the DC-part of $R^T x(t) x^T(t) R$, which is $Cov_v(0)$.

The settling of this oscillation towards the stationary state decreases as ρ increases, but at the same time the amplitude of the oscillation increases. This causes the instability of the overall system for large ρ. For sufficiently small amplitudes, i.e. for small $\rho > 0$ a similar behaviour as for (4.1) can be expected.

5 Simulation Results

In order to demonstrate the convergence of the algorithms and the typical form of results obtained in practice, consider the problem of teaching a network to replicate the simple illustrative signal

$$r(t) = \sin t + \sin 2t \tag{1}$$

consisting of $N = 2$ distinct frequency components of roughly equal significance. The network chosen consists of the model in $\mathbb{R}^5$ with network parameter $\alpha = 1$:

$$\dot{x}(t) = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 0 \\ w_2 & w_2 & w_3 & w_4 & -1 \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} w(t) \tag{2}$$

$$y(t) = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{3}$$

The initial state $x(0)$ is specified by $x(0) = 0$.

The subject of this simulation is the application of the base algorithm (2.12) on the above example. It shows the influence of the learning rate parameter Q on the speed of convergence and demonstrates exponential convergence of the net output and of the weights towards the ideal weights. The result of a simulation of the learning phase with $Q = 100\ I$ are shown in figures 1 a-c.

(i) Practical convergence is achieved at the output in effectively $t = 400(s)$ (i.e. in 130 periods of r) (Fig. 1a) and of the weights in $t = 700(s)$ (Fig. 1b).

(ii) The linear decrease of the output and the parameter error e_1 and e_p in a logarithmic scale substantiates the analytical result of exponential convergence given in theorem 3.6 (Fig. 1c). Note that the errors stop decreasing at $t = 1600(s)$ due only to the fixed precision of the numerical calculation.

The result of a simulation of the learning phase with $Q = 1$ is shown in figure 2a: Again the convergence is exponential, however, the rate of convergence is distinctly less in this case, hence confirming Q as a parameter for influencing convergence rates in a systematic way.

In figure 2b-c the weight errors of the original system $e_{p_i}(t)$ and the averaged system $e_{pav_i}(t)$, $i = 1, \ldots, 4$ are compared. The simulations demonstrate for $Q = I$ the good approximation of the original system by the averaged system. This is confirmed by the following table which contains γ_{av} for $Q = \epsilon I$, $\epsilon = 1, 100$, and the true exponential convergence rate γ determined numerically from the simulation data.

ϵ	γ	γ_{av}
1	$4.4628 \cdot 10^{-4}$	$4.9511 \cdot 10^{-4}$
100	$7.9551 \cdot 10^{-3}$	$4.9511 \cdot 10^{-2}$

Furthermore the adaptive computation of Q based on formula (4.1) and (4.2) leads to the expected acceleration of the convergence. This is demonstrated by the simulations in figure 2d-e. For $\epsilon = 1$ the accelerated algorithm practical convergence is achieved at the output already in $t = 25(s)$ (i.e. 4 perods of r) and for the weights in $t = 40(s)$ (Figure 2d). Figure 2e shows in comparison with figure 1c the substantial acceleration of the exponential error decrease.

For a further simulation another reference signal consisting of $N = 5$ distinct frequency components was selected:

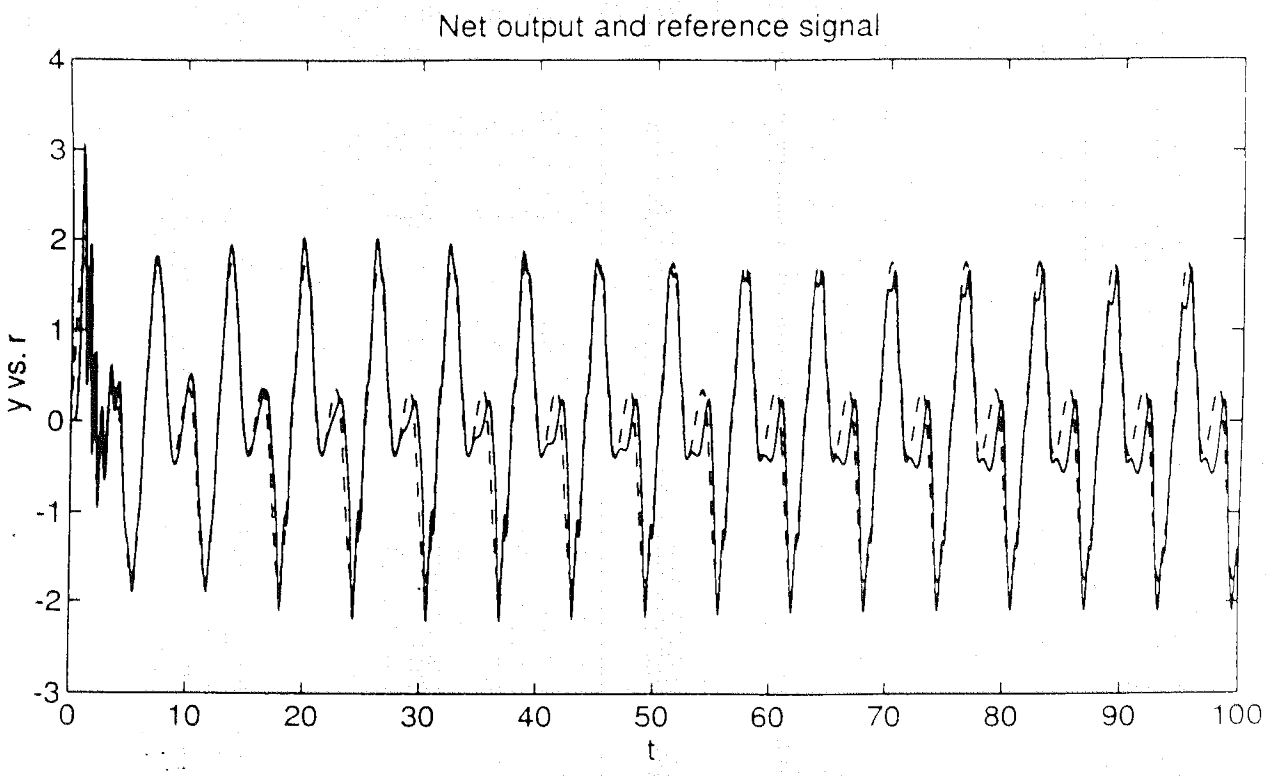
$$r(t) = \sum_{w=1}^{5} \sin wt$$

The network chosen consists of a minimal net of dimension $n = 11$. The corresponding ideal weight vector is

$$w^* = (-442000, 749000, -759400, 458524, -216260, 67650, -18260, 2827, -440, 0)^T.$$

All initial states are again 0. The smallest real part of the eigenvalues of $\mathrm{Cov}_{v,\bar{v}}(0)$ in this example is $3.8854 \cdot 10^{-13}$. The averaging method predicts for $Q = \epsilon \cdot I$ an estimation of the necessary length of the learning phase of $t^u = \frac{\log(10^6)}{3.8854 \cdot 10^{-13}}(s)$ to achieve a relative precision of 10^{-6} in the weights. This gives $t^u \approx 3.6 \cdot 10^{11}(s)$ for $Q = 100I$. Obviously a simulation of the learning phase doesn't make sense in this case.

However, the adaptive Liapunov learning rule (4.1) and (4.2) provides for the above example an estimation of the exponential convergence rate $\gamma \approx \gamma_{av} = \frac{1}{26} = 0.0385$ and $t^u \approx \frac{\log(10^6)}{\gamma_{av}} \approx 3.45(s)$ for the length of the learning phase for $\epsilon = 1$. The simulations confirm this estimations. Figures 3a-c contain the simulation results for the above example with $\rho = 0.04$ and $\epsilon = 1$.
Practical convergence is achieved at the output in effectively $t = 700(s)$ (i.e. in 110 periods of r) and of the weights in $t = 800(s)$ (Fig. 3b). Opposite to the case of 2 frequencies the simulations show a transient phase of $600(s)$ before rapid exponential convergence is achieved. The length of this second phase of about 300(s) is in good accordance with the above estimation of the length of the learning phase. This observation is quantitatively confirmed by the estimated rate $\gamma_{av} = 0.0385$ and the actual exponential convergence rate computed from the simulation data as $\gamma = 0.0414$.



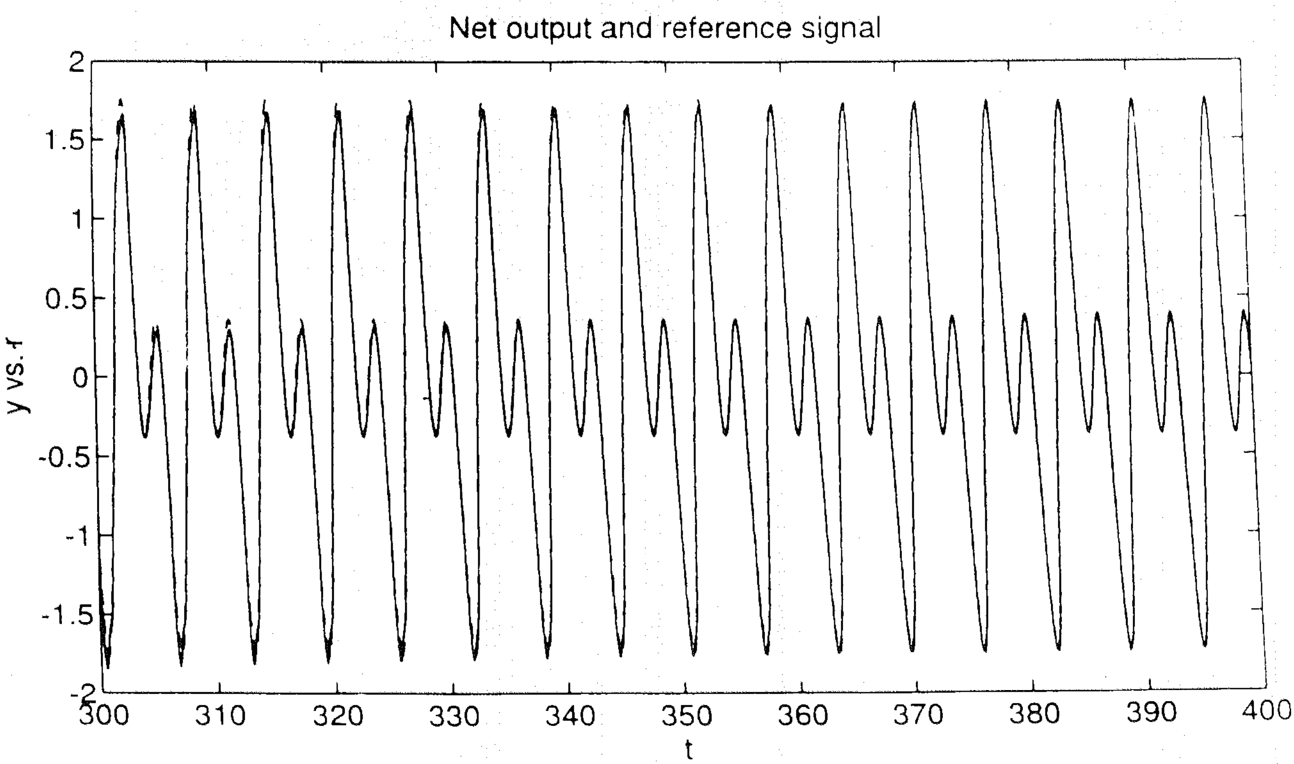
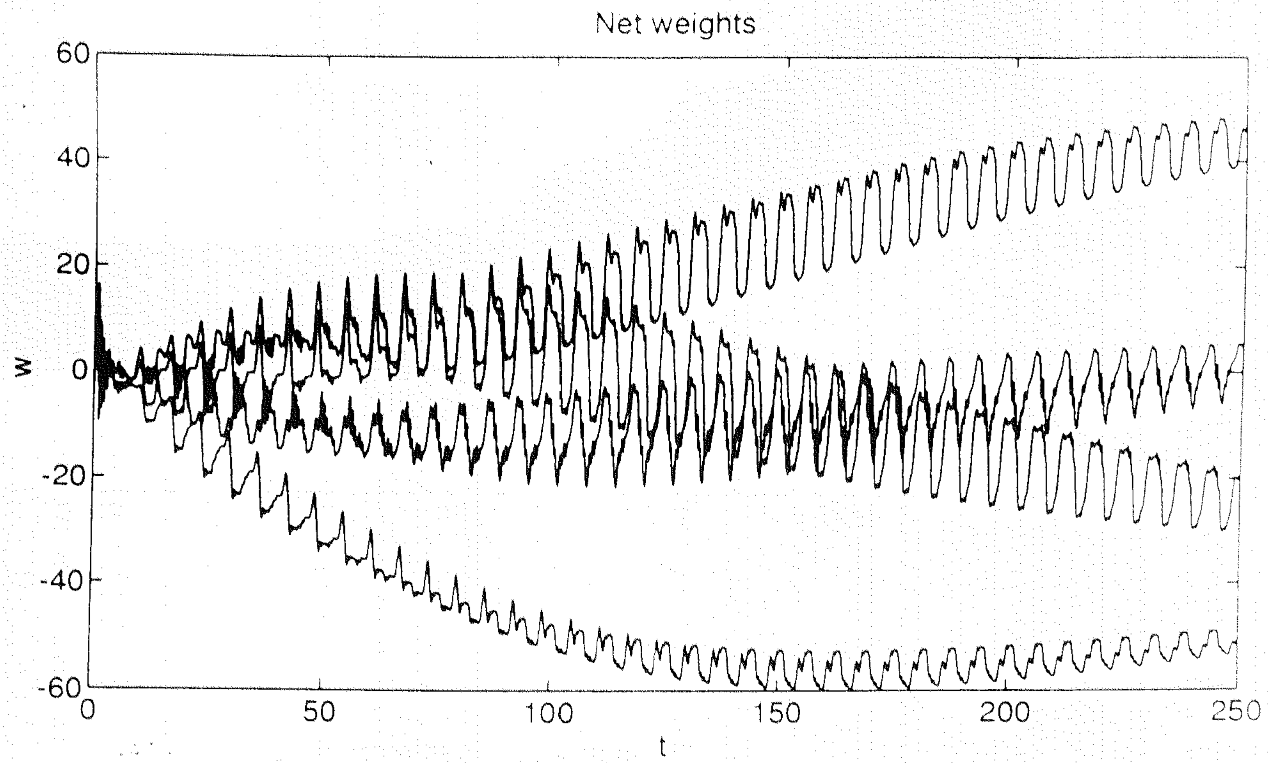


Figure 1a: Basic learning rule with $Q = 100I$

— $y(t)$, - - - $r(t)$



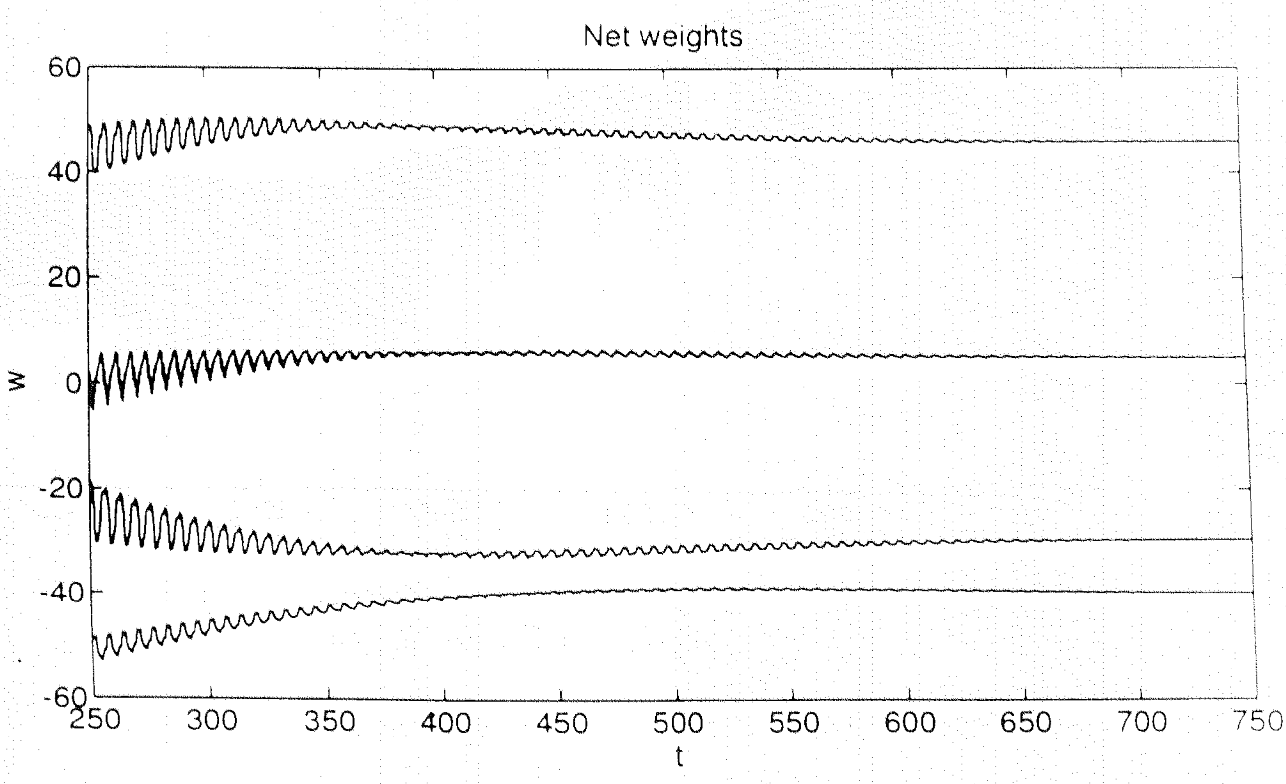


Figure 1b: Basic learning rule with $Q = 100I$
$w^* = (-40, 46, -30, 5)^T$

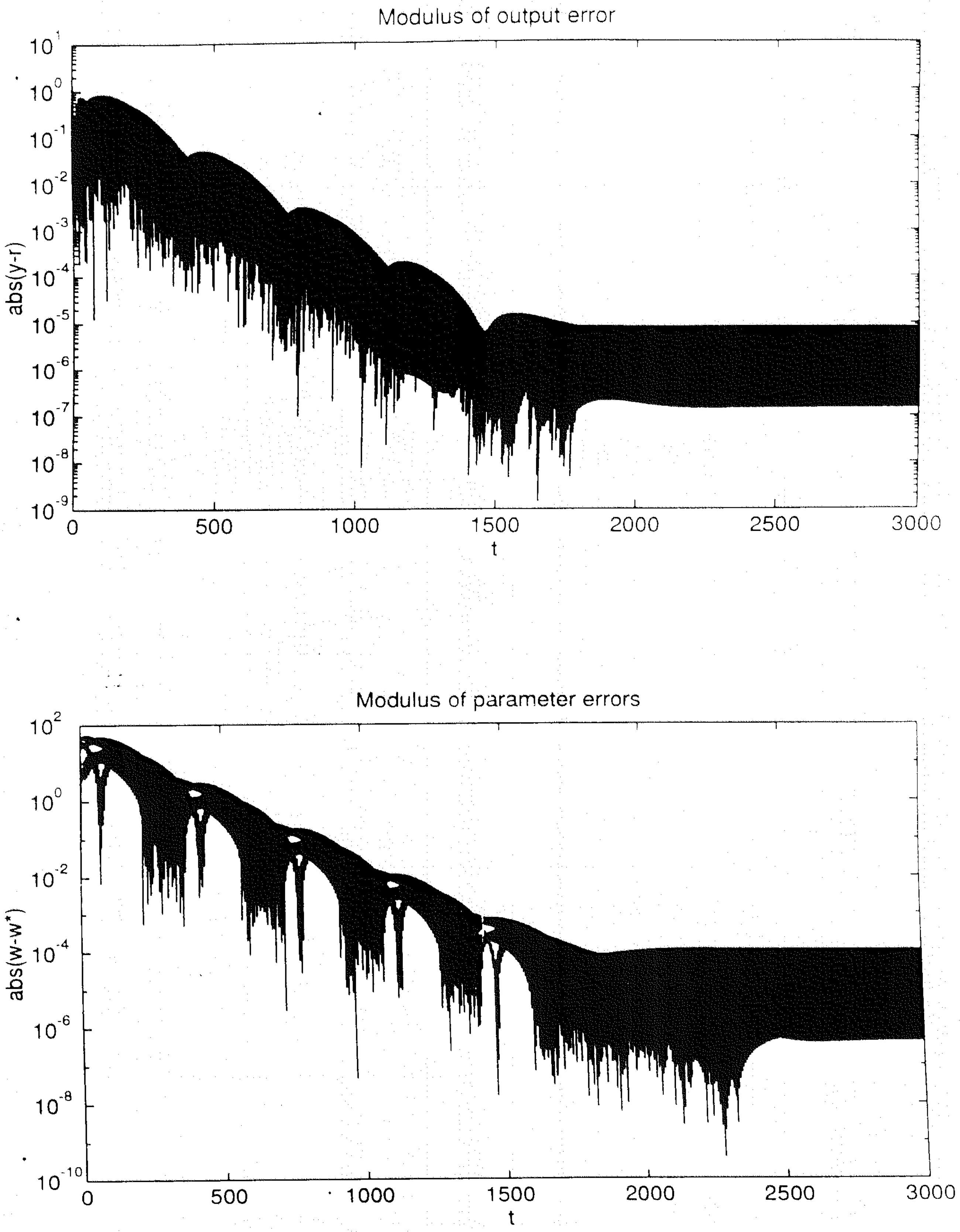


Figure 1c: Basic learning rule with $Q = 100I$
above: $|e_1(t)|$, below: $|e_{p_1}(t)|$

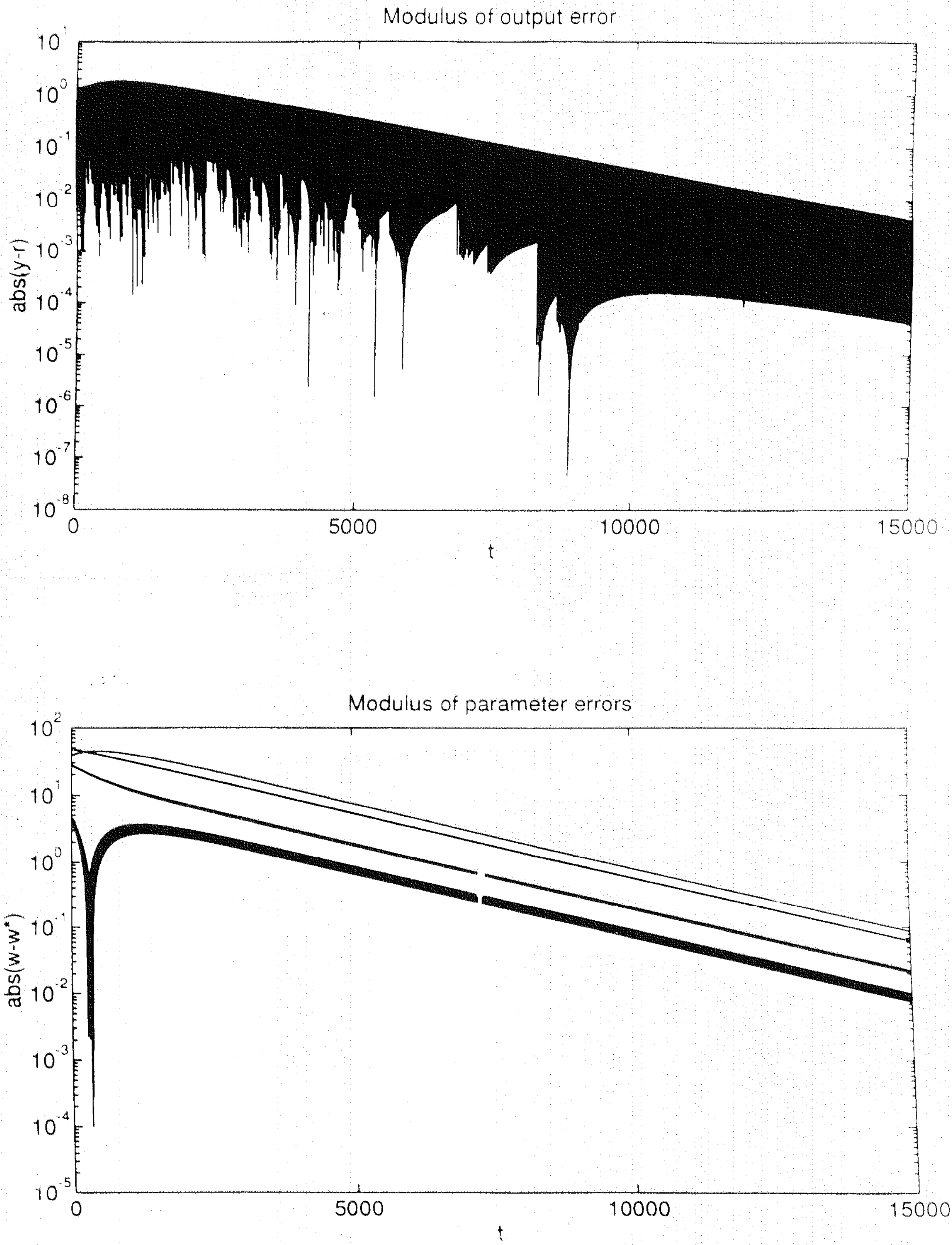


Figure 2a: Basic learning rule with $Q = 1$
above: $|e_1(t)|$, below: $|e_{p_1}(t)|$

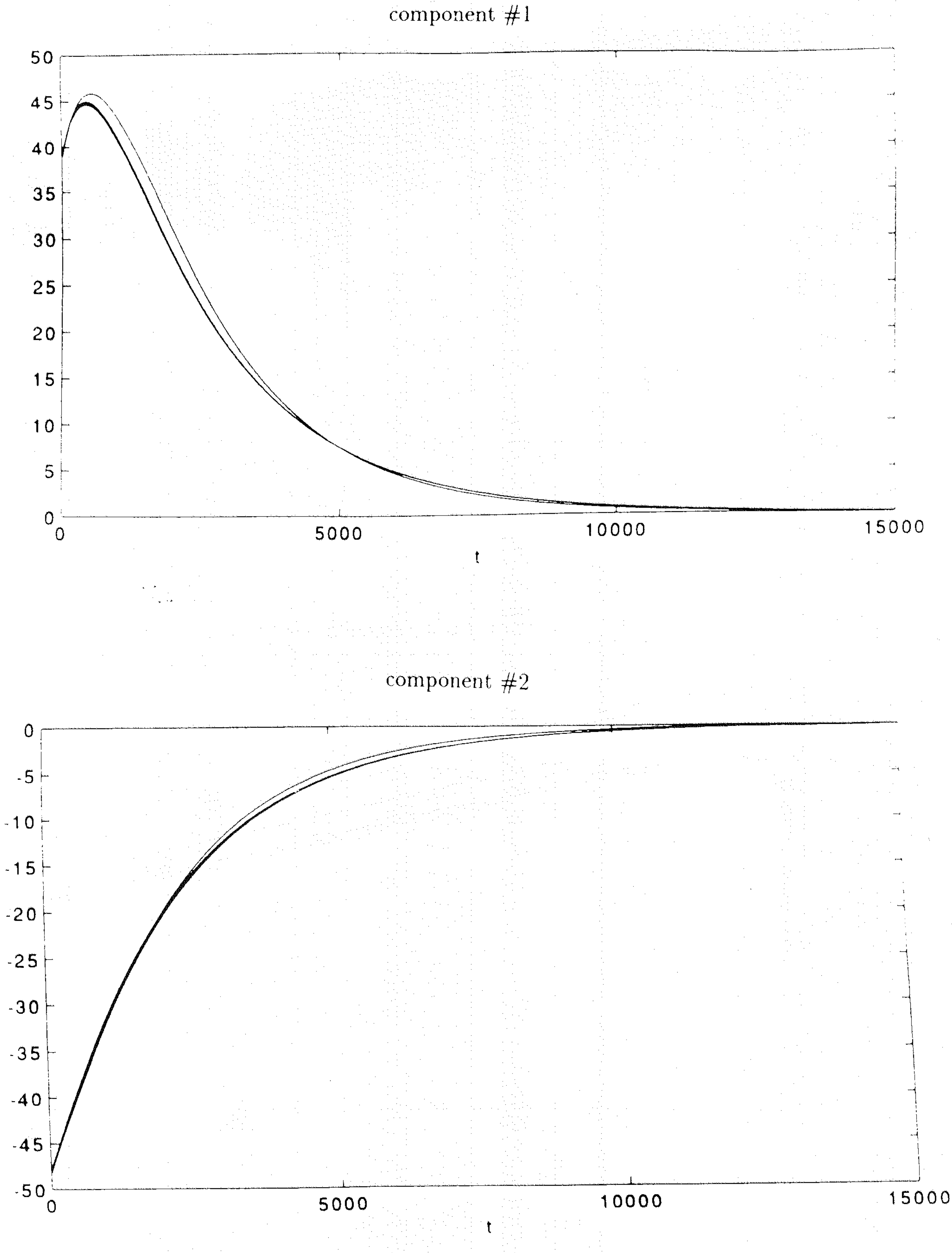
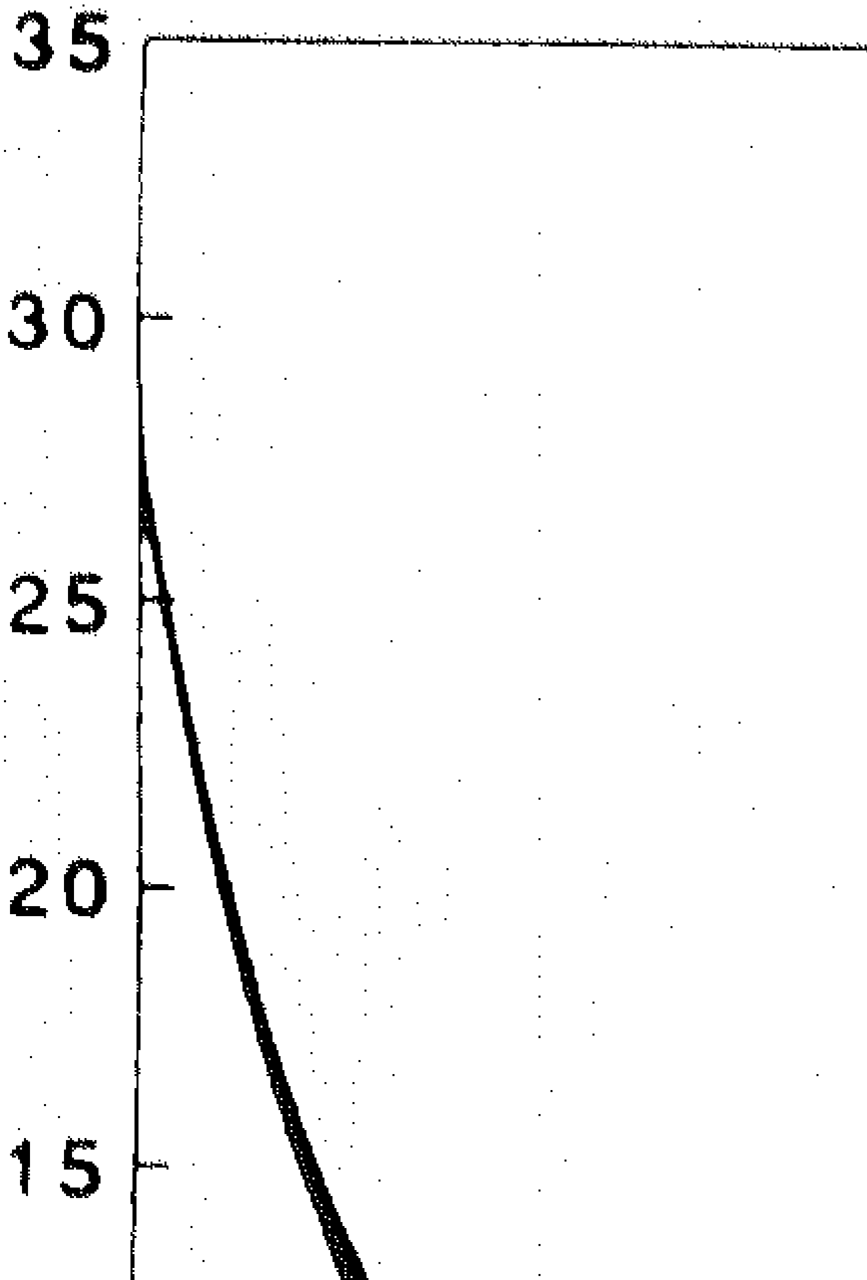
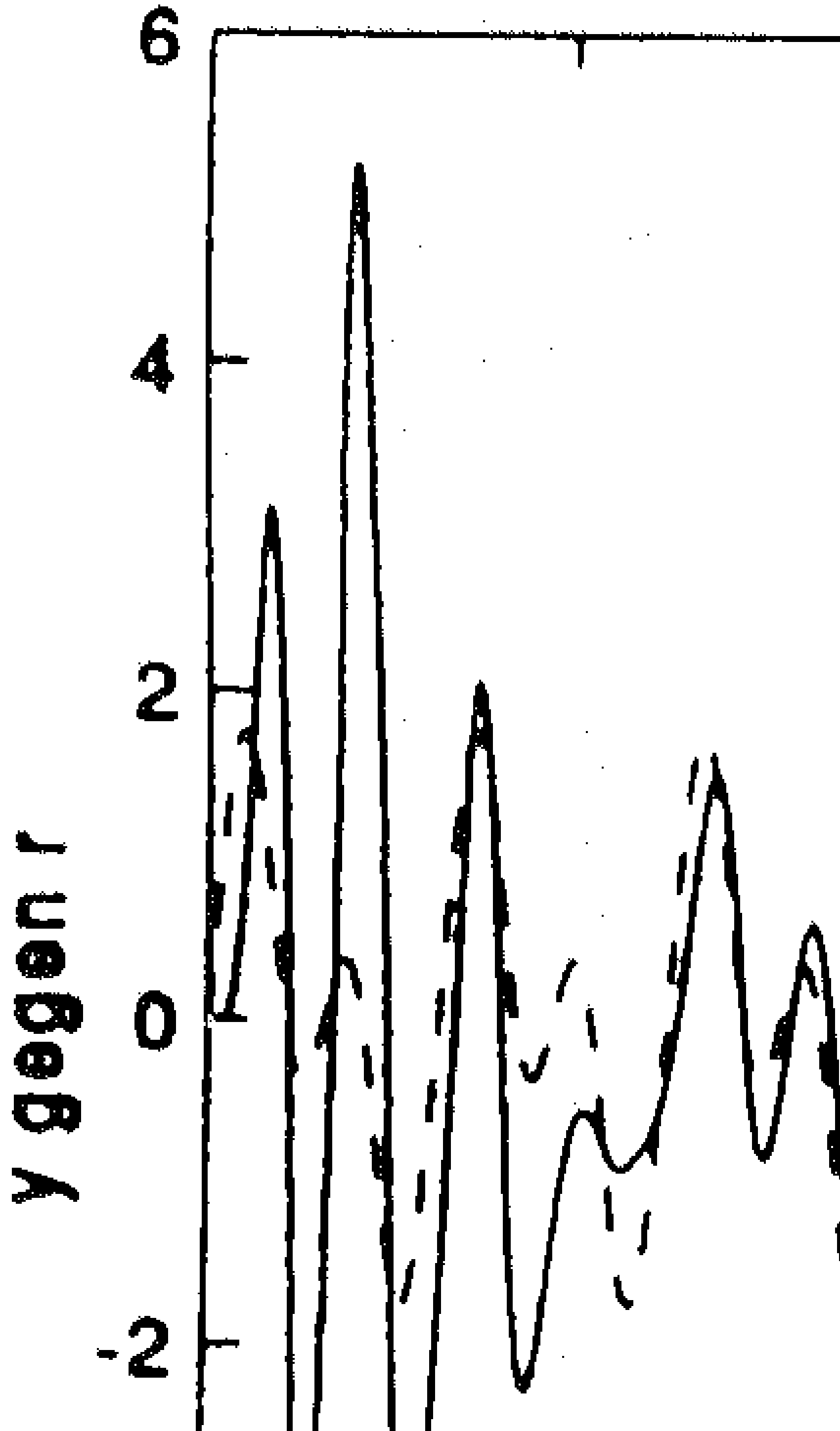
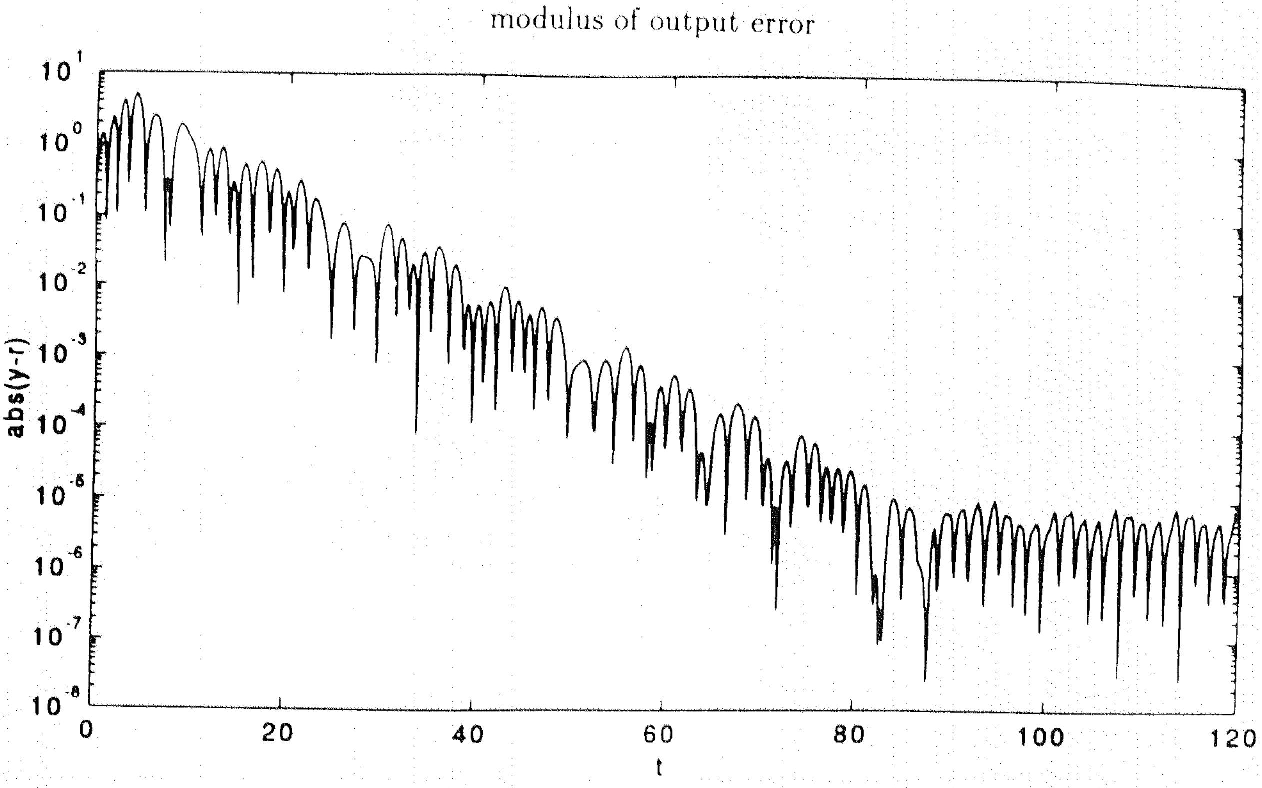


Figure 2b: Basic learning rule with $Q = I$, weight errors for the original system $e_{p_1}(t)$ (fat) and averaged system $e_{p_{av_1}}(t)$


35
30
25
20
15



6
4
2
0
-2
y gegen r



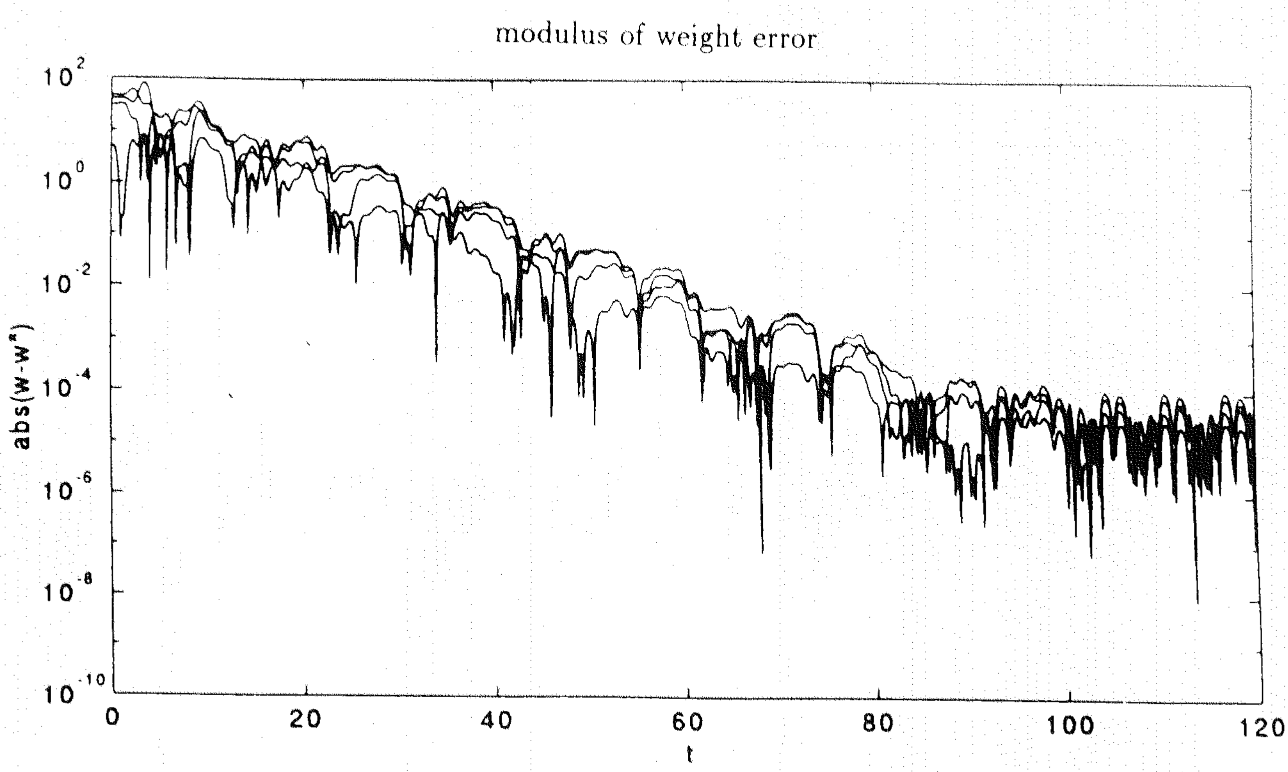
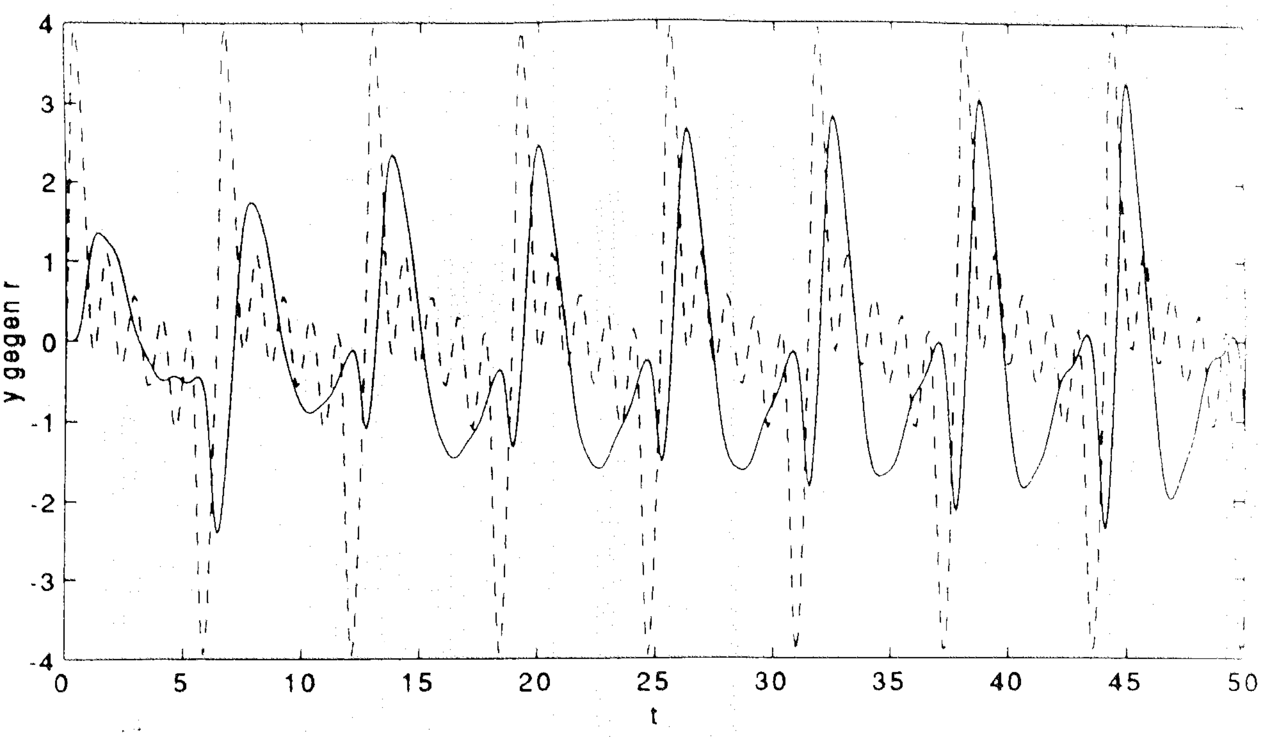


Figure 2e: Learning rule with adaptive Q.
above: $|e_1(t)|$, below: $|e_{p_i}(t)|$

net output and reference signal



net output and reference signal

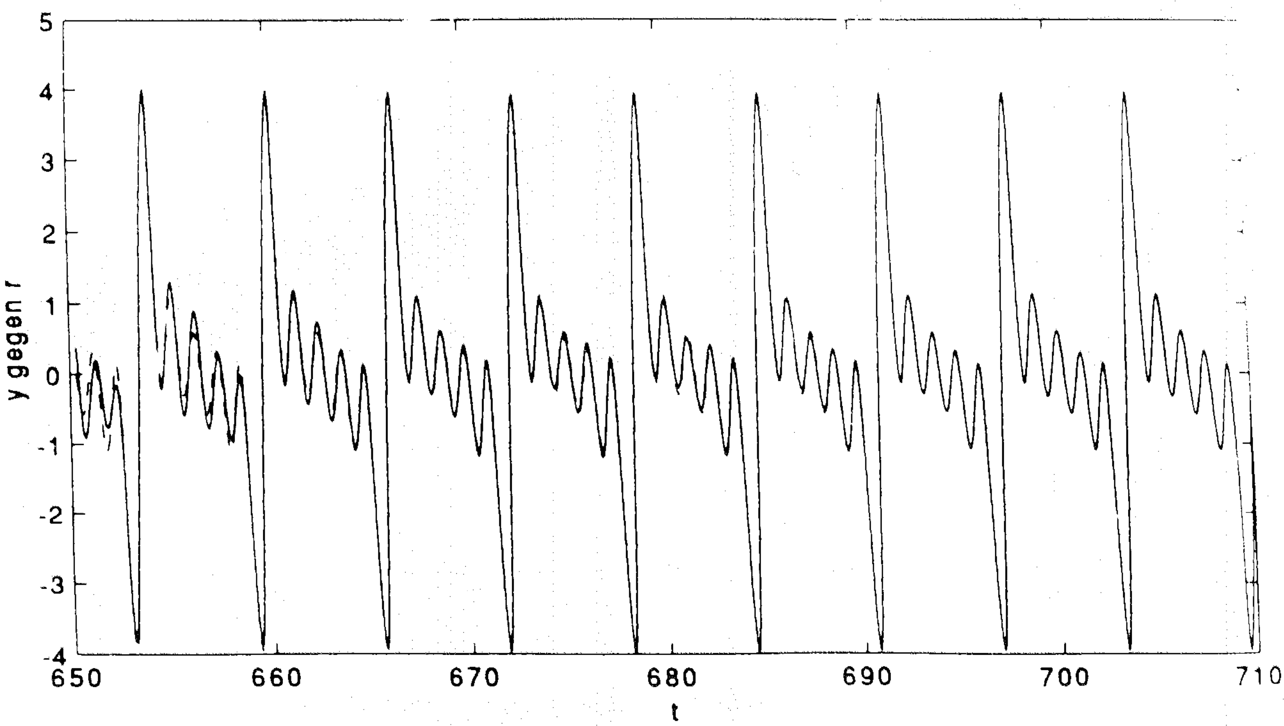


Figure 3a: Learning rule with adaptive Q.
-y(t),- - - r(t)

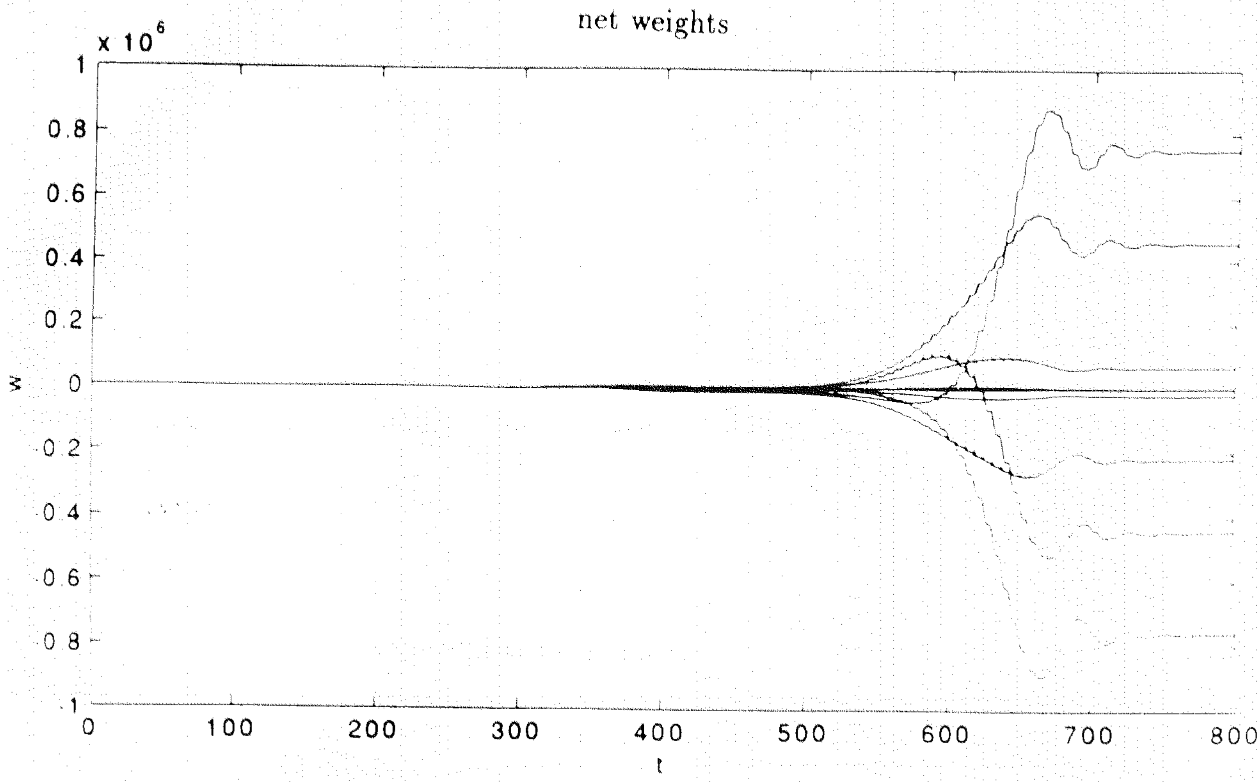
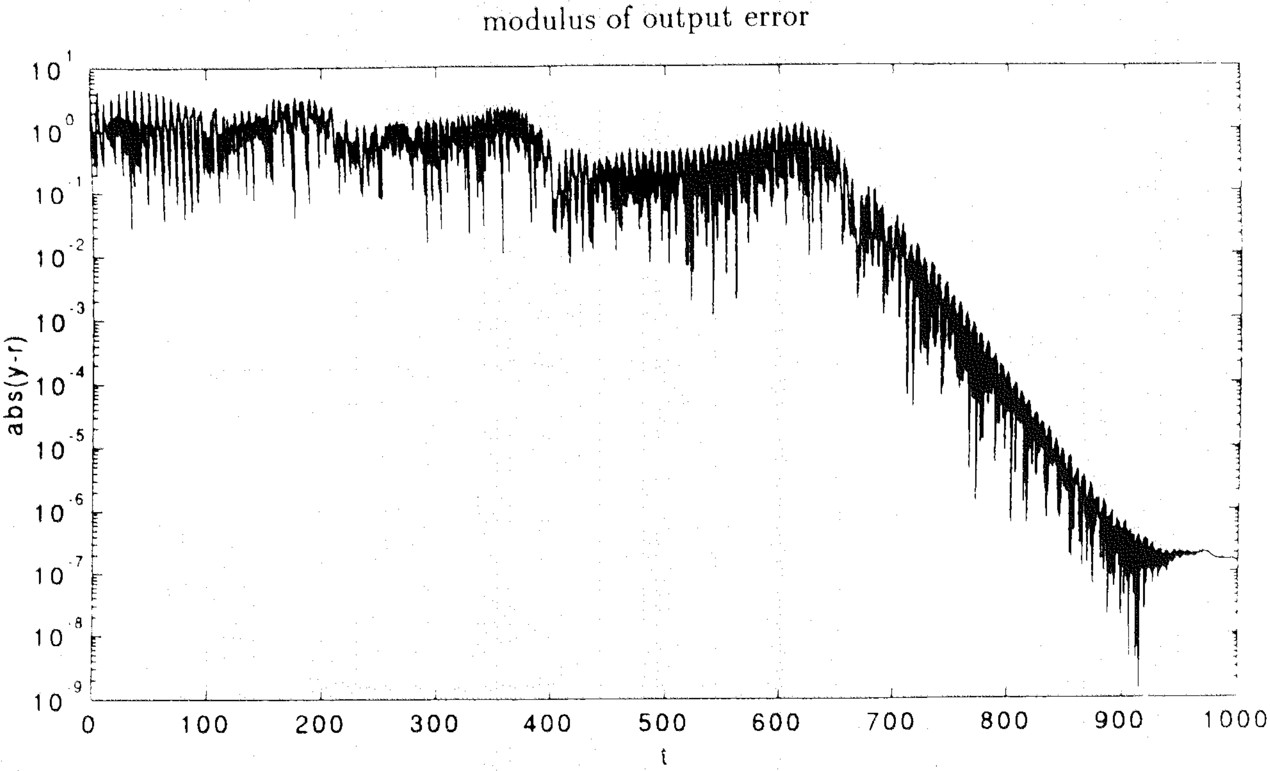


Figure 3b: Learning rule with adaptive Q.

$w^* = (-442000, 749000, -759400, 458524, -216260, 67650, -18260, 2827, -440, 0)^T$



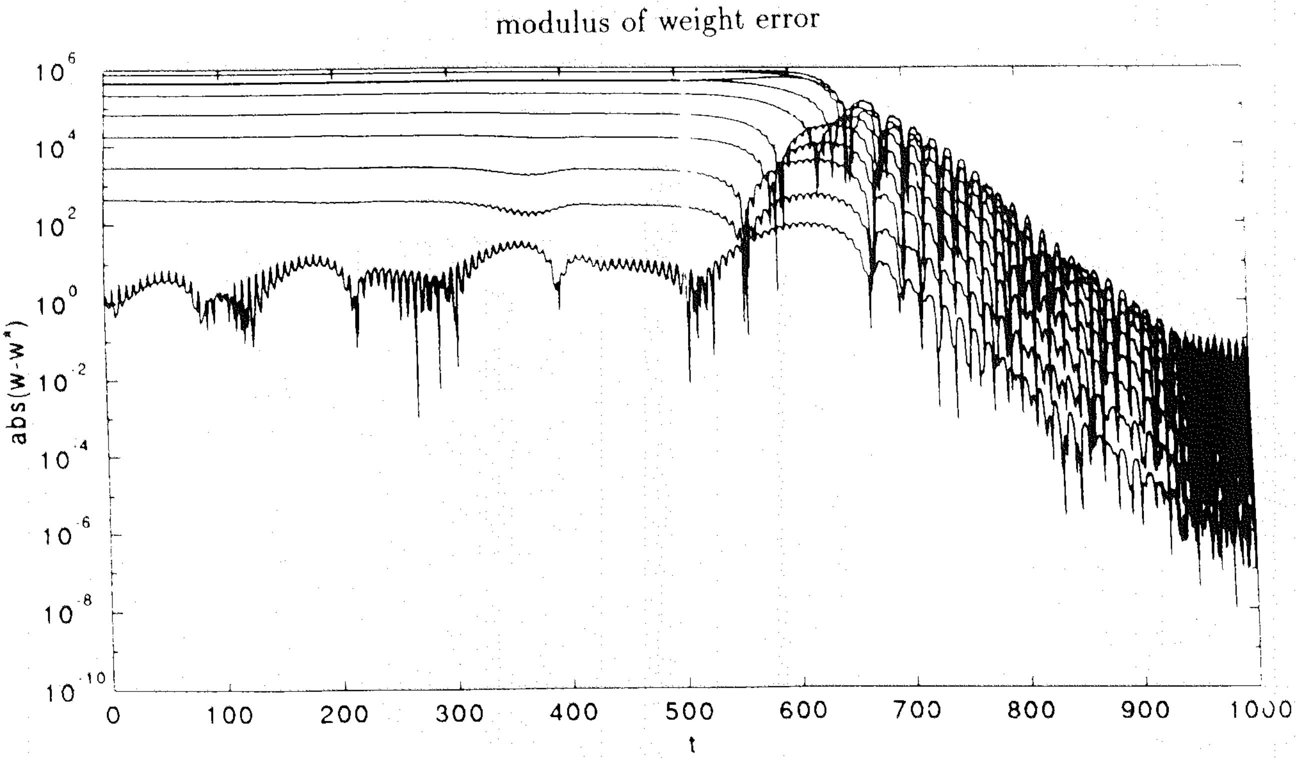


Figure 3c: Learning rule with adaptive Q.

6. References

[1] Arnold, V. *Geometric methods in the theory of ordinary differential equations.* Springer, Berlin, 1982.

[2] Balachandra, M. and P.R. Sethna. *A generalization of the method of averaging for systems with two-time scales.* Archive for Rational Mechanics and Analysis, pp. 261-283, 1975.

[3] Bochner, S. *Vorlesungen über Fouriersche Ingetrale.* Akademische Verlagsgesellschaft, Leipzig, 1932. Nachdruck, Chelsea Publ., 1948.

[4] Bogoliuboff, N.N. and Y.A. Mitropolskii. *Asymptotic methods in the theory of nonlinear oscillators.* Gordon & Breach, New York, 1961.

[5] Boyd, S. and S. Sastry. *On parameter convergence in adaptive control.* System Control letters, 3: 311-319, 1983.

[6] Doya, K. and S. Yoshisawa. *Adaptive neural oscillator using continuous-time back-propagation learning.* Neural Networks, 2: 375-385, 1989.

[7] Fu L. C., M. Bodson and S. Sastry. *New stability theorems for averaging and application to convergence analysis of adaptive identification and control schemes.* In P. Kokotovic, B. Bensoussan and G. Blankenship (Eds.), Singular Perturbations and Asymptotic Analysis in Control Systems. Lecture Notes in Control and Information Sciences, Springer, New York, 1986.

[8] Guckenheimer J. and P. Holmes. *Nonlinear oscillations, dynamical systems and bifurcations of vector fields.* Springer, Berlin, 1967.

[9] Hale, J.K. *Ordinary differential equations.* Krieger, Huntington, New York, 1980.

[10] Reinke, R., D. Prätzel-Wolters and D.H. Owens. *Learning and replication of periodic signals in dynamic networks.* Proceedings of the third European Control Conference ECC 95, Rome 1995, pp. 2118-2125.

[11] Reinke, R. *Adaptive Regeln zum Lernen und Reproduzieren periodischer Signale mit dynamischen Netzwerken.* Ph. D. thesis, University of Kaiserslautern, Germany, 1994.

[12] Sanders, J.A. and F. Verhulst. *Averaging methods in nonlinear dynamical systems.* Springer, Berlin, 1985.

[13] Sethna, P.R. *Method of averaging for systems bounded for positive time.* Journal of Math. Anal. and Applications, 41:621-631, 1973.

[14] Volosov, V.M. *Averaging in systems of ordinary differential equations.* Russian Mathematical Surveys, 17(6): 1-126, 1962.