**Proceedings of the 29th Conference
on Decision and Control
Honolulu, Hawaii • December 1990**

FP-7-2 - 4:40

Learning Control for a Closed Loop System using *Feedback-Error-Learning*

Hiroaki GOMI Mitsuo KAWATO
ATR Auditory and Visual Perception Research Laboratories,
Seika-cho, Soraku-gun, Kyoto 619-02, Japan

Abstract

In this paper, we propose new learning schemes using *feedback-error-learning* for a neural network model applied to adaptive nonlinear feedback control. Using these schemes, the actual responses after learning correspond to desired responses. When the desired response in Cartesian space is required, learning impedance control is derived. The convergence properties of the neural networks are provided by averaged equation and Liapunov method. We show also some simulation results of these learning schemes.

1. Introduction

Adaptive control theory has mainly been applied to linear systems. As for applications for nonlinear robot systems, Dubowsky [1] applied Model Reference Adaptive Control (MRAC) to a robot manipulator control using the local linearization technique, and Slotine [2] proposed adaptive control method based on the Liapunov function for a manipulator whose nonlinear characteristics were known in advance.

In previous studies of adaptive learning control using a neural network model, Barto [3], Jordan [4] and Psaltis [5] addressed the problem of how to obtain the error signal for the neural network controller. The difference between the reference response and the actual response cannot directly be used as the error for controller adaptation because of the controlled object characteristics between the controller output (i.e., input of the controlled object) and actual response (i.e., the output of controlled object). Thus, Barto proposed using ASE and ACE techniques [3], and Jordan proposed using forward and inverse modeling [4], to solve this problem.

Kawato et al. proposed a feedforward controller learning scheme, which uses a feedback signal as the error for training a neural network model [6]. They call this learning method *feedback-error-learning*. Using this method, the neural network model for feedforward compensation acquires an inverse dynamics model of the controlled object. This method was originally proposed as a model of voluntary movement learning in animals. However a feedforward controller doesn't improve the controllability of the feedback system.

In this paper we propose learning schemes using *feedback-error-learning* for a Neural network model applied to an adaptive Nonlinear Feedback Controller (NNFC). The essential point of these learning schemes is that a Conventional Feedback Controller(CFC) is provided both as a usual feedback controller to guarantee global asymptotic stability and as an inverse reference model of the response of the controlled object. If the reference model of the response in Cartesian space is prepared as the CFC, *impedance control* [7] is derived after learning. Additionally, convergence properties of the neural networks are shown by using an averaged equation method and the Liapunov's second method. We also pointed out the relationship of these learning schemes to the cerebellum's posture and locomotion adaptive control mechanism in animals.

2. Adaptive nonlinear feedback controller

In this section, we propose two adaptive learning control methods using *feedback-error-learning* technique for neural network feedback controllers.

In the first learning scheme, the neural network feedback controller finally acquires an inverse dynamics model of the controlled object. Thus, we call this learning scheme Inverse Dynamics Model Learning (IDML). In the second learning scheme, the neural network model is trained to become a nonlinear regulator to compensate for the nonlinearity (except for the inertia term) of the controlled object through learning. Accordingly, this is called Nonlinear Regulator Learning (NRL). These two learning schemes are described in detail below. The convergence properties of the neural networks of these two learning schemes are derived by using the similar procedure of the proof for stability as the *feedback-error-learning* for feedforward adaptive control provided by Kawato [8].

2.1 Inverse Dynamics Model Learning(IDML)

We explain the configuration of this learning system, and provide the proof of the convergence property of this learning scheme. Subsequently, we show the learning impedance control as an application of IDML.

The conventional feedback controller (CFC) is used both as a ordinary feedback controller to guarantee global asymptotic stability during the learning period, and as a reference model of the response of the controlled object. The components, CFC, NNFC and Controlled object, are connected as shown in Fig.1.

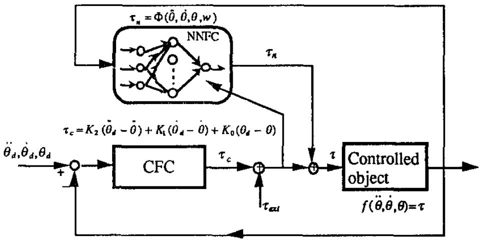

$\tau_n = \Phi(\ddot{\theta}, \dot{\theta}, \theta, w)$
NNFC
τ_n
$\tau_c = K_2(\ddot{\theta}_d - \ddot{\theta}) + K_1(\dot{\theta}_d - \dot{\theta}) + K_0(\theta_d - \theta)$
$\ddot{\theta}_d, \dot{\theta}_d, \theta_d$
+
CFC
τ_c
τ
Controlled object
τ_{ext}
$f(\ddot{\theta}, \dot{\theta}, \theta) = \tau$

Fig. 1 Inverse Dynamics Model Learning

The sum of the output of the CFC and the disturbance for a controlled object is fed to the NNFC as the error signal. This corresponds to the synaptic modification input. The neural network feedback controller also receives $\theta, \dot{\theta}$ and $\ddot{\theta}$ as the usual synaptic inputs. As the neural network acquires the inverse dynamics model of the controlled object through learning, the response of the controlled object follows the desired response implemented as the inverse reference model in the conventional feedback controller.

We here express each component and this learning scheme in the equation form for the proof of convergence. The controlled object dynamics *f* are expressed as:

$$f(\ddot{\theta}, \dot{\theta}, \theta) = \tau \tag{1}$$

The equation describing the response of the CFC is linear as given below in Eq. 2 in this explanation.

$$\tau_c = K_2(\ddot{\theta}_d - \ddot{\theta}) + K_1(\dot{\theta}_d - \dot{\theta}) + K_0(\theta_d - \theta) \tag{2}$$

NNFC is expressed as:

$$\tau_n = \Phi(\ddot{\theta}, \dot{\theta}, \theta, w) \tag{3}$$

Here θ is the actual position vector, θ_d is the desired position vector and w is the synaptic weight (adaptive parameter) of NNFC. Any neural network model that has the capacity to approximate to nonlinear objective function by minimizing error signal for the NNFC, can be arranged as the NNFC. For example, Multi Layer Perceptron (MLP) [9], Cerebellar Model Articulator Controller (CMAC) [10], Table Look Up, and Memory Based Reasoning (MBR) [11] may be used as NNFC. The equation of the system is

$$\tau = \tau_n + \tau_c + \tau_{ext} \tag{4}$$

The difference of τ and τ_n is expressed as τ_{imag}.

$$\tau_{imag} = \tau - \tau_n = \tau_c + \tau_{ext} \tag{5}$$

Eq. 5 can be rewritten in Eq. 6 using Eq. 2.

$$K_2(\ddot{\theta} - \ddot{\theta}_d) + K_1(\dot{\theta} - \dot{\theta}_d) + K_0(\theta - \theta_d) = \tau_{ext} - \tau_{imag} \tag{6}$$

The learning rule of the *feedback-error-learning* scheme is represented in Eq. 7 in a general manner in this learning scheme.

$$\frac{dw}{dt}=\eta\left(\frac{\partial\Phi}{\partial w}\right)^T(\tau_c+\tau_{ext})=\eta\left(\frac{\partial\Phi}{\partial w}\right)^T\tau_{imag} \qquad (7)$$

where η is a positive definite matrix which determines learning rate.

The convergence property of the NNFC is shown below by using the technique of averaging random differential equation, Geman's theorem and the Liapunov's second method under following three assumptions.

Assumption 1

The learning rate, η, is very small and positive definite.

Assumption 2

The inputs, τ_{ext} and $\ddot{\theta}_d,\dot{\theta}_d,\theta_d$, are strongly mixing and strongly stationary stochastic processes.

Assumption 3

The CFC is designed to guarantee for the convergence of θ before learning, through learning and also after learning.

We can express the system dynamics in following random differential equations instead of the deterministic equation, Eq.4 and Eq.7, under these assumptions.

$$\begin{aligned}f(\ddot{\theta}(t,\omega),\dot{\theta}(t,\omega),\theta(t,\omega))-\Phi(\ddot{\theta}(t,\omega),\dot{\theta}(t,\omega),\theta(t,\omega),w(t,\omega))\\=K_2\left(\ddot{\theta}_d(t,\omega)-\ddot{\theta}(t,\omega)\right)+K_1\left(\dot{\theta}_d(t,\omega)-\dot{\theta}(t,\omega)\right)\\+K_0\left(\theta_d(t,\omega)-\theta(t,\omega)\right)+\tau_{ext}(t,\omega),\end{aligned} \qquad (8)$$

$$\begin{aligned}\frac{dw(t,\omega)}{dt}=\eta\left(\frac{\partial\Phi(\ddot{\theta}(t,\omega),\dot{\theta}(t,\omega),\theta(t,\omega),w(t,\omega))}{\partial w}\right)^T\\\times\left\{\tau_c(\ddot{\theta}(t,\omega),\dot{\theta}(t,\omega),\theta(t,\omega))+\tau_{ext}(t,\omega)\right\},\end{aligned} \qquad (9)$$

here ω is a sample point in a probability space. The solution of the following averaged equation is a good approximate to the solution of Eq.9 for small η, as proven by Geman's theorem [12].

$$\begin{aligned}\frac{dM}{dt}=\eta E\left[\left(\frac{\partial\Phi(\ddot{\theta}(t,\omega),\dot{\theta}(t,\omega),\theta(t,\omega),w(t,\omega))}{\partial w}\right)^T\right.\\\left.\times\left\{\tau_c(\ddot{\theta}(t,\omega),\dot{\theta}(t,\omega),\theta(t,\omega))+\tau_{ext}(t,\omega)\right\}\right]_{w=M}\end{aligned} \qquad (10)$$

Thus, we consider the following function V as a candidate of the Liapunov function for the averaged equation (10) by using following function L.

$$\begin{aligned}L=\frac{1}{2}\left\{\tau_c(\ddot{\theta}(t,\omega),\dot{\theta}(t,\omega),\theta(t,\omega))+\tau_{ext}(t,\omega)\right\}^T\\\times\left\{\tau_c(\ddot{\theta}(t,\omega),\dot{\theta}(t,\omega),\theta(t,\omega))+\tau_{ext}(t,\omega)\right\}\end{aligned} \qquad (11)$$

$$\begin{aligned}&V(M,t)\\&=E\left[L(\ddot{\theta}(t,\omega),\dot{\theta}(t,\omega),\theta(t,\omega),\tau_{ext}(t,\omega))\right]_{w=M}\\&=E\left[L(\ddot{\theta}(t,\omega),\dot{\theta}(t,\omega),\theta(t,\omega),\ddot{\theta}_d(t,\omega),\dot{\theta}_d(t,\omega),\theta_d(t,\omega),\right.\\&\qquad\left.w(t,\omega),\tau_{ext}(t,\omega))\right]_{w=M}\ \geq 0 \qquad (\because \text{Eq.5})\end{aligned} \qquad (12)$$

The time derivative of V is calculated as follows.

$$\begin{aligned}\frac{dV(M,t)}{dt}=E\left[\left\{\tau_c(\ddot{\theta}(t,\omega),\dot{\theta}(t,\omega),\theta(t,\omega))+\tau_{ext}(t,\omega)\right\}^T\right.\\\left.\times\frac{d}{dt}\left\{\tau_c(\ddot{\theta}(t,\omega),\dot{\theta}(t,\omega),\theta(t,\omega))+\tau_{ext}(t,\omega)\right\}\right]_{w=M}\end{aligned} \qquad (13)$$

In order to simplify the expression, we use shortened expression for each variable.

$$\frac{dV}{dt}=E\left[(\tau_c+\tau_{ext})^T\left\{\frac{\partial\tau_c}{\partial\ddot{\theta}}\frac{d\ddot{\theta}}{dt}+\frac{\partial\tau_c}{\partial\dot{\theta}}\frac{d\dot{\theta}}{dt}+\frac{\partial\tau_c}{\partial\theta}\frac{d\theta}{dt}+\frac{d\tau_{ext}}{dt}\right\}\right]_{w=M} \qquad (14)$$

Eq.8 can be expressed in an explicit form with respect to $\ddot{\theta}$ as follows:

$$\ddot{\theta}(t,\omega)=h(\dot{\theta}(t,\omega),\theta(t,\omega),w(t,\omega),\ddot{\theta}_d(t,\omega),\dot{\theta}_d(t,\omega),\theta_d(t,\omega),\tau_{ext}(t,\omega)). \qquad (15)$$

By differentiating Eq.15 with respect to t, and then substituting it for $d\ddot{\theta}/dt$ in Eq.14, following equation is obtained.

$$\begin{aligned}\frac{dV}{dt}=E\left[(\tau_c+\tau_{ext})^T\left\{-K_2\left(\frac{\partial\ddot{\theta}}{\partial w}\frac{dw}{dt}+\frac{\partial\ddot{\theta}}{\partial\dot{\theta}}\frac{d\dot{\theta}}{dt}+\frac{\partial\ddot{\theta}}{\partial\theta}\frac{d\theta}{dt}+\frac{\partial\ddot{\theta}}{\partial\ddot{\theta}_d}\frac{d\ddot{\theta}_d}{dt}\right.\right.\right.\\\left.\left.\left.+\frac{\partial\ddot{\theta}}{\partial\dot{\theta}_d}\frac{d\dot{\theta}_d}{dt}+\frac{\partial\ddot{\theta}}{\partial\theta_d}\frac{d\theta_d}{dt}+\frac{\partial\ddot{\theta}}{\partial\tau_{ext}}\frac{d\tau_{ext}}{dt}\right)-K_1\frac{d\dot{\theta}}{dt}-K_0\frac{d\theta}{dt}+\frac{d\tau_{ext}}{dt}\right\}\right]_{w=M}\end{aligned} \qquad (16)$$

We group the above terms into two parts with respect to w and θ in order to eliminate terms of zero by averaging.

$$\begin{aligned}\frac{dV}{dt}=E\left[(\tau_c+\tau_{ext})^T\left\{-K_2\frac{\partial\ddot{\theta}}{\partial w}\right\}\frac{dw}{dt}\right]_{w=M}\\+E\left[(\tau_c+\tau_{ext})^T\left\{-K_2\left(\frac{\partial\ddot{\theta}}{\partial\dot{\theta}}\frac{d\dot{\theta}}{dt}+\frac{\partial\ddot{\theta}}{\partial\theta}\frac{d\theta}{dt}+\frac{\partial\ddot{\theta}}{\partial\ddot{\theta}_d}\frac{d\ddot{\theta}_d}{dt}\right.\right.\right.\\\left.\left.\left.+\frac{\partial\ddot{\theta}}{\partial\dot{\theta}_d}\frac{d\dot{\theta}_d}{dt}+\frac{\partial\ddot{\theta}}{\partial\theta_d}\frac{d\theta_d}{dt}+\frac{\partial\ddot{\theta}}{\partial\tau_{ext}}\frac{d\tau_{ext}}{dt}\right)-K_1\frac{d\dot{\theta}}{dt}-K_0\frac{d\theta}{dt}+\frac{d\tau_{ext}}{dt}\right\}\right]_{w=M}\end{aligned} \qquad (17)$$

Then the second term of Eq.17 vanishes, if θ,θ_d,τ_d are strongly stationary stochastic processes. The reason is as follows. The second term of Eq.17 can be expressed as:

$$\begin{aligned}\frac{d}{dt}E\left[L(\ddot{\theta}(t,\omega),\dot{\theta}(t,\omega),\theta(t,\omega),\ddot{\theta}_d(t,\omega),\dot{\theta}_d(t,\omega),\theta_d(t,\omega),\right.\\\left.M,\tau_{ext}(t,\omega))\right]\end{aligned} \qquad (18)$$

Because of the assumption 3, the differential equation of θ is stable. Hence, $\ddot{\theta},\dot{\theta},\theta$ and L become Baire functions of the strongly stationary processes, $\ddot{\theta}_d,\dot{\theta}_d,\theta_d$ and τ_{ext}. Consequently, $E\left[L(\ddot{\theta},\dot{\theta},\theta,\ddot{\theta}_d,\dot{\theta}_d,\theta_d,M,\tau_{ext})\right]$ doesn't depend on time and the second therm vanishes. Hence, we obtain:

$$\frac{dV}{dt}=E\left[(\tau_c+\tau_{ext})^T\left\{-K_2\frac{\partial\ddot{\theta}}{\partial w}\right\}\frac{dw}{dt}\right]_{w=M}. \qquad (19)$$

We calculate partial derivative of Eq.15 with respect to w while referring to Eq.8 in order to replace $\partial\ddot{\theta}/\partial w$ in Eq.19.

$$\begin{aligned}&\frac{\partial f}{\partial\ddot{\theta}}\frac{\partial\ddot{\theta}}{\partial w}-\frac{\partial\Phi}{\partial\ddot{\theta}}\frac{\partial\ddot{\theta}}{\partial w}-\frac{\partial\Phi}{\partial w}=-K_2\frac{\partial\ddot{\theta}}{\partial w}\\&\Leftrightarrow\quad\frac{\partial\ddot{\theta}}{\partial w}=\left(\frac{\partial f}{\partial\ddot{\theta}}-\frac{\partial\Phi}{\partial\ddot{\theta}}+K_2\right)^{-1}\frac{\partial\Phi}{\partial w}\end{aligned} \qquad (20)$$

Then, Eq.19 is expressed in:

$$\begin{aligned}\frac{dV}{dt}&=E\left[(\tau_c+\tau_{ext})^T\left\{-K_2\left(\frac{\partial f}{\partial\ddot{\theta}}-\frac{\partial\Phi}{\partial\ddot{\theta}}+K_2\right)^{-1}\frac{\partial\Phi}{\partial w}\right\}\frac{dw}{dt}\right]_{w=M}\\&=E\left[-(\tau_c+\tau_{ext})^TK_2\left(\frac{\partial f}{\partial\ddot{\theta}}-\frac{\partial\Phi}{\partial\ddot{\theta}}+K_2\right)^{-1}\frac{\partial\Phi}{\partial w}\eta\left(\frac{\partial\Phi}{\partial w}\right)^T(\tau_c+\tau_{ext})\right]_{w=M}.\end{aligned} \qquad (21)$$

Considering the assumption 3, the following matrix becomes positive definite.

$$K_2\left(\frac{\partial f}{\partial\ddot{\theta}}-\frac{\partial\Phi}{\partial\ddot{\theta}}+K_2\right)^{-1} \qquad (22)$$

Therefore, the middle part of Eq.21 is positive semi-definite as follows.

$$K_2\left(\frac{\partial f}{\partial\ddot{\theta}}-\frac{\partial\Phi}{\partial\ddot{\theta}}+K_2\right)^{-1}\frac{\partial\Phi}{\partial w}\eta\left(\frac{\partial\Phi}{\partial w}\right)^T \qquad (23)$$

Thus, we obtain:

$$\frac{dV}{dt}\leq 0. \qquad (24)$$

The equalities of Eq.12 and 24 hold only when $w=\hat{w}$ for non trivial θ except for the local minimum condition. Consequently, we conclude that w asymptotically converges to the optimal set of synaptic weights, $\hat{w}$, if the conditions of local minimum are avoided through learning.

We describe above the learning scheme in Joint space, but it also can be applied in Cartesian space. We explain briefly the learning impedance control, as an application of the IDML, by using the CFC as the task-oriented reference model in Cartesian space. The manipulator dynamics can be expressed in Eq.25 and CFC is represented in Eq.26 by using the Cartesian coordinate.

$$R(\theta)\ddot{\theta}+N(\theta,\dot{\theta})=\tau \qquad (25)$$

$$F_c=M\ddot{x}_e+B\dot{x}_e+Kx_e \qquad (26)$$

where $x_e=x_d-x$

And the function Φ of the neural network is the same equation as Eq.3. Total system is designed as shown in Fig. 2.

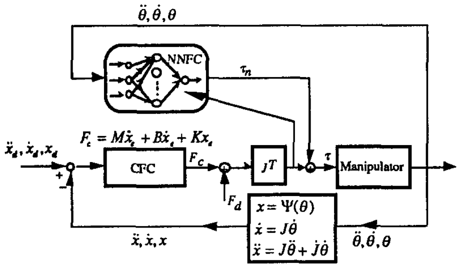

$\ddot{\theta},\dot{\theta},\theta$
NNFC
τ_n
$\ddot{x}_d,\dot{x}_d,x_d$
$F_c = M\ddot{x}_e + B\dot{x}_e + Kx_e$
CFC
F_c
J^T
τ
Manipulator
+
-
F_d
$x = \Psi(\theta)$
$\dot{x} = J\dot{\theta}$
$\ddot{x} = J\ddot{\theta} + \dot{J}\dot{\theta}$
$\ddot{x},\dot{x},x$
$\ddot{\theta},\dot{\theta},\theta$

Fig.2 Learning Impedance Control

By using the same procedure as shown in former proof, the convergence property of this learning is obtained. Consequently, the total feedback compensation, $\tau_n + \tau_c$, is

$$\tau_n + \tau_c = R(\theta)\ddot{\theta} + N(\theta,\dot{\theta}) - J^T\left(M(\ddot{x}-\ddot{x}_d) + B(\dot{x}-\dot{x}_d) + K(x-x_d)\right). \quad (27)$$

Hence, the response of the end-point of the manipulator is:

$$M(\ddot{x}-\ddot{x}_d) + B(\dot{x}-\dot{x}_d) + K(x-x_d) = F_{ext}. \quad (28)$$

Thus, the manipulator is operated by *impedance control* [7] in Cartesian space after the learning is completed. Moreover, we can change the virtual impedance of the manipulator into any value by simply changing the parameters *M,B,K* of the CFC without relearning, because the NNFC obtains only the inverse dynamics model of the controlled object.

2.2 Nonlinear Regulator Learning (NRL)

We explain here the configuration of this learning system and the consequence after learning, and then show the convergence property briefly.

The dynamics of the controlled object is represented here as:

$$R(\theta)\ddot{\theta} + N(\theta,\dot{\theta}) = \tau \quad (29)$$

As with the first method, the CFC serves the same two purposes (see 2.1). However we set only the output of the CFC, τ_c, as the error signal for the NNFC as shown in Fig.3. In this learning scheme, the disturbance is considered to be absent in the learning period.

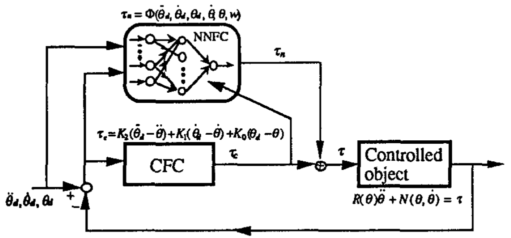

$\tau_n = \Phi(\ddot{\theta}_d, \dot{\theta}_d, \theta_d, \dot{\theta}, \theta, w)$
NNFC
τ_n
$\tau_c = K_2(\ddot{\theta}_d - \ddot{\theta}) + K_1(\dot{\theta}_d - \dot{\theta}) + K_0(\theta_d - \theta)$
CFC
τ_c
τ
Controlled object
$R(\theta)\ddot{\theta} + N(\theta,\dot{\theta}) = \tau$
$\ddot{\theta}_d, \dot{\theta}_d, \theta_d$
+
-

Fig. 3 Nonlinear Regulator Learning

In this case, we feed the desired values, $\ddot{\theta}_d, \dot{\theta}_d, \theta_d$, and the outputs of object , $\theta, \dot{\theta}$, to the NNFC. Hence, the function of the NNFC is

$$\tau_n = \Phi\left(\ddot{\theta}_d, \dot{\theta}_d, \theta_d, \dot{\theta}, \theta, w\right). \quad (30)$$

The total dynamics is given by Eq.31.

$$\begin{aligned} &\tau = \tau_c + \tau_n \\ &\Leftrightarrow \left(R(\theta) + K_2\right)(\ddot{\theta}_d - \ddot{\theta}) + K_1(\dot{\theta}_d - \dot{\theta}) \\ &\qquad + K_0(\theta_d - \theta) + \Phi - N(\theta,\dot{\theta}) - R(\theta)\ddot{\theta}_d = 0 \end{aligned} \quad (31)$$

If Φ is represented as in Eq.32, Eq.31 is expressed as in Eq.33.

$$\Phi_d = N(\theta,\dot{\theta}) + R(\theta)\ddot{\theta}_d + K_2^{-1}R(\theta)\left(K_1(\dot{\theta}_d - \dot{\theta}) + K_0(\theta_d - \theta)\right) \quad (32)$$

$$\begin{aligned} &\left(R(\theta) + K_2\right)(\ddot{\theta}_d - \ddot{\theta}) + K_1(\dot{\theta}_d - \dot{\theta}) + K_0(\theta_d - \theta) \\ &\qquad + K_2^{-1}R(\theta)\left(K_1(\dot{\theta}_d - \dot{\theta}) + K_0(\theta_d - \theta)\right) = 0 \end{aligned} \quad (33)$$

When K_2 is selected so as to satisfy Eq.34, Eq.35 is obtained.

$$K_2 R(\theta) = R(\theta) K_2 \quad (34)$$

$$\left(R(\theta) + K_2\right)\left(K_2(\ddot{\theta}_d - \ddot{\theta}) + K_1(\dot{\theta}_d - \dot{\theta}) + K_0(\theta_d - \theta)\right) = 0 \quad (35)$$

Consequently, this equation yields the free response of the controlled objects as:

$$K_2\ddot{\xi} + K_1\dot{\xi} + K_0\xi = 0. \quad (36)$$

Here $\xi = \theta_d - \theta$, and under the condition $R(\theta) + K_2 \neq 0$.

Roughly speaking, the neural network model is trained to become a nonlinear regulator to obtain the reference response in the free movement by compensation for the nonlinearity (except for the inertia term) of the controlled object.

The convergence property of the neural network in this learning scheme is explained below shortly. The difference of the output of the NNFC and desired function is defined as P. P is represented in Eq. 37 by using Eq. 29, 31, 32.

$$P = \Phi - \Phi_d = -\left(I + K_2^{-1}R(\theta)\right)\tau_c \qquad I\text{: unit matrix} \quad (37)$$

We consider the following function J as a candidate of the Liapunov function of this learning scheme.

$$J = E\left[\frac{1}{2}P^T P\right]_{w=M} \geq 0 \quad (38)$$

Using the same procedure of the former proof of convergence property, the following equation is obtained.

$$\frac{dJ}{dt} \leq 0 \quad (39)$$

Consequently, the function Φ of the neural network acquires nonlinear compensation represented in Eq.32 after learning.

As the IDML method, we can also provide this learning scheme in Cartesian space. But impedance control is not derived perfectly because the NNFC doesn't have the feedback acceleration signal, $\ddot{\theta}$. Therefore, inertia term of virtual impedance cannot make small by using the NRL method in Cartesian space.

2.3 Discussion

Comparing the consequence of IDML and NRL, there is an advantage of IDML because the nonlinearity of inertia term of controlled object is compensated by using feedback acceleration signal for the NNFC. Same responses in the free movement are obtained after learning by using each learning scheme, but the NRL method cannot provide a linear response for the disturbance, τ_{ex}, when there is the nonlinearity of the inertia term of the controlled object.

That is because that the dynamics of controlled object is not fully compensated for by using feedback signals in the NRL method. However, as the disturbance is not used as the error signal for training of the NNFC in NRL, we need not measure or estimate the disturbance in learning. The choice of learning schemes depend on the controlled object and practical usage.

As for the problem of the local minimum of learning a neural network, there is no method to perfectly solve or avoid it. But there are recently many good learning examples by using neural network [9], so we think that the solution obtained by using the IDML or NRL method is in the neighborhood of the global minimum.

3. Simulation results

In this section, we show the simulation results of the learning schemes described in the previous section, applied to some control problems. The inverted pendulum is used as a controlled object for IDML and NRL in the joint space. Additionally a two-link manipulator is used as a controlled object for IDML in Cartesian space (i.e., learning impedance control).

3.1 Inverted Pendulum Controlled by IDML

The dynamics of the controlled object, an inverted pendulum, is expressed in Eq.40. The input u in this equation is the acceleration of a cart, the second term in the left side of Eq.40 correspond to Coulomb's friction. This is a typical regulator problem because the desired values $\ddot{\theta}_d, \dot{\theta}_d, \theta_d$ are always zero.

$$-\left(J + ml^2\right)\ddot{\theta} - \left(\frac{2}{1+\exp(-a\dot{\theta})} - 1\right) + mgl\sin\theta = mlu\cos\theta \quad (40)$$

here	m : pendulum weight	l : pendulum length
	J : pendulum inertia	θ : angle of pendulum
	u : input accel.	g : accel. of gravity

The following prepared nonlinearities, p_i, that are needed for making the inverse dynamics model of this controlled object, are used as the input of the two-layered neural network in the NNFC.

$$p_0 = \tan\theta$$
$$p_1 = \left(2/\left(1+\exp(-a\dot{\theta})\right)-1\right)/\cos\theta$$
$$p_2 = \ddot{\theta}/\cos\theta \qquad (41)$$

The output of NNFC τ_n is

$$\tau_n = \sum_{k=0}^{2} w_k p_k \qquad (42)$$

If Eq. 40 is the objective function for the NNFC, the desired value of the weights are

$$w_0 = g,\quad w_1 = -\frac{1}{ml},\quad w_2 = -\frac{J+ml^2}{ml} \qquad (43)$$

Using these prepared nonlinearities and linear parameter adaptation technique, the function, $(\partial\Phi/\partial w)(\partial\Phi/\partial w)^T$, is positive definite except when the condition, $\theta=\dot{\theta}=\ddot{\theta}=0$, is satisfied. Thus, the Liapunov function V is strictly concave because τ_{imag} is not zero except for the condition, $\theta=\dot{\theta}=\ddot{\theta}=0$. Hence, in this case, it is guaranteed that the global optimal synaptic weights are obtained by the *feedback-error-learning* rule.

First, the time courses of weight values during IDML is shown in Fig.4. We used Ornstein-Uhlenbeck Process as the input disturbance in this learning period. $\hat{w}_2, \hat{w}_1, \hat{w}_0$ are the desired values for the actual values, w_2, w_1, w_0, respectively. Each actual value approached to each desired value gradually. Fig.5 shows the moving average of square of the NNFC output and the error for the NNFC during the IDML period. The NNFC output increased for compensation of controlled object dynamics. Such adaptation caused the error for the NNFC to decreased.

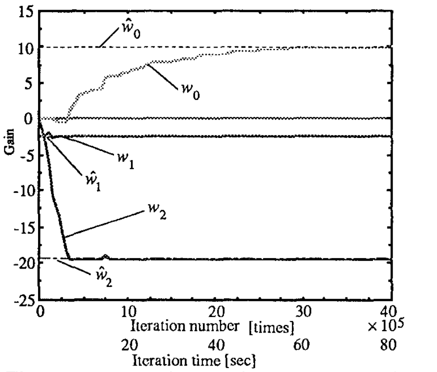


Fig. 4 Time courses of weights during inverse dynamics model learning

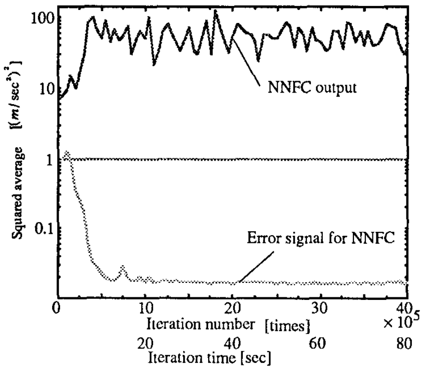


Fig. 5 Moving average of square of the NNFC output and the error for NNFC during inverse dynamics model learning

The responses of the pendulum for the input disturbance are compared in Fig.6. The response after learning corresponded perfectly to the desired response such as:

$$K_2(\ddot{\theta}-\ddot{\theta}_d)+K_1(\dot{\theta}-\dot{\theta}_d)+K_0(\theta-\theta_d)=\tau_{ext} \qquad (44)$$

(In this case, $\ddot{\theta}_d=\dot{\theta}_d=\theta_d=0$, $K_2=1, K_1=3.5, K_0=20$)

The linear controller designed to get the desired response for a linearized controlled object without the friction term such as:

$$-\left(J+ml^2\right)\ddot{\theta}+mgl\theta = mlu \qquad (45)$$

The response driven by only the linear controller did not corresponded to the desired response because of object nonlinearity and unconsidered factor, friction.

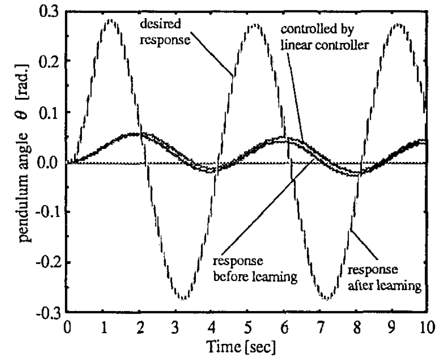


Fig. 6 Controlled object response for disturbance $\tau_{ext} = -5\sin(\pi t/2)[m/\sec^2]$

3.2 Inverted Pendulum Controlled by NRL

We show here one result of the second learning scheme, NRL, using the same controlled object, inverted pendulum. We have already confirmed that the actual weight values approached the desired weight values also in this learning scheme by using prepared nonlinearities, so following result is derived by using general neural network (i.e., 3-layered and sigmoid function in hidden layer unit) to confirm the ability of this learning scheme.

Fig.7 shows the desired response, the response using the linear controller, the response before learning and the response after learning. The linear controller is designed to get the desired response for the controlled object expressed in Eq.45. We did not get the same response by using a linear controller with desired response because of the nonlinearity and friction of the actual controlled object. As the nonlinear adaptive controller compensated for the nonlinearity of the controlled object through learning, the actual response approached the desired response.

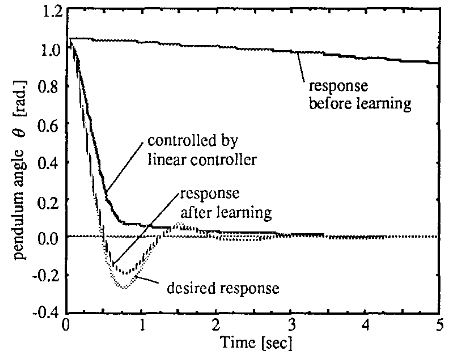


Fig. 7 Controlled object response from initial position of 60[deg] $\tau_c = -\left[1.0(\ddot{\theta}-\ddot{\theta}_d)+3.5(\dot{\theta}-\dot{\theta}_d)+20.0(\theta-\theta_d)\right]$

3.3 Learning Impedance Control of the 2-link manipulator

In this section, we show some results of learning impedance control using the 2-link manipulator.

Fig.8 shows the responses of the end-point of the manipulator in each condition (before learning, after learning, ideal response). If decoupled impedance is required for the manipulator, the ideal response for the step force input along x or y axis, becomes the straight path along each axis as shown in Fig. 8 under the perfect impedance control conditions. The responses after learning were more contiguous to the ideal response than before learning. Thus, we can conclude that the NNFC acquired an approximate model of the inverse dynamics model in a work space which was explored during learning.

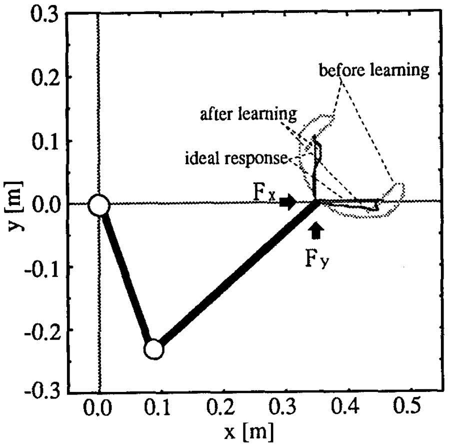


Fig. 8 Improved response for the force inputs at the end effector
$F_c = -[0.1(\ddot{x}-\ddot{x}_d)+2.0(\dot{x}-\dot{x}_d)+20.0(x-x_d)]$

Next results are the responses of the contact task for the wall (which is modeled as a strong spring). Before learning, the actual trajectory differed from the desired trajectory in free movement, and contact task was still unstable as shown in Fig.9 and 10. After learning, the actual trajectory was much better as shown in Fig.11, and the impact force at the time when collision occurred, was much lower than before learning as shown in Fig.12, because virtual impedance of the manipulator was lowered. Moreover, the contact task was so stable that the external force was in proportion to magnitude of the difference of actual trajectory and desired trajectory as shown in Fig.12, because virtual impedance was changed to get stable condition in contact phase. The parameters of the CFC are shown below in each condition.

$F_c = -[M\ddot{x}+B\dot{x}+K(x-x_d)]$ $\quad x=[x,y]^T$

Before learning	M=*diag*[0.1, 0.1][kg]
	B=*diag*[7.0, 7.0][N/(m/sec)]
	K=*diag*[100, 100][N/m]
After learning	M=*diag*[0.1, 0.1][kg]
	B=*diag*[7.0, 7.0][N/(m/sec)] (uncontact phase)
	B=*diag*[700, 7.0][N/(m/sec)](contact phase)
	K=*diag*[100, 100][N/m]

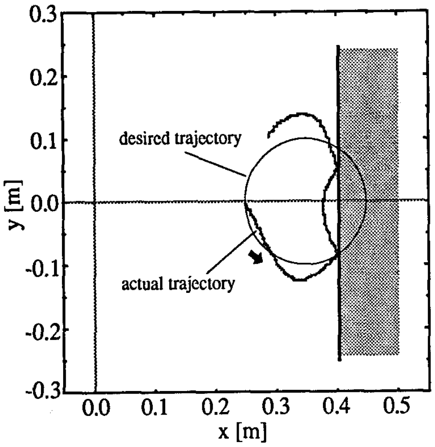


Fig. 9 Trajectory of free movement & Contact task **before learning**

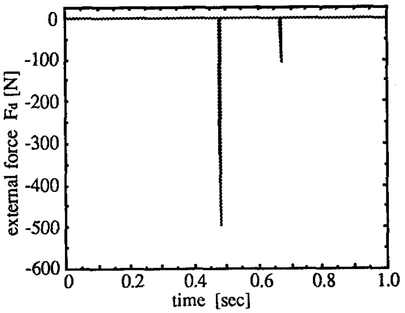


Fig. 10 External force at the end effector of free movement & Contact task **before learning**

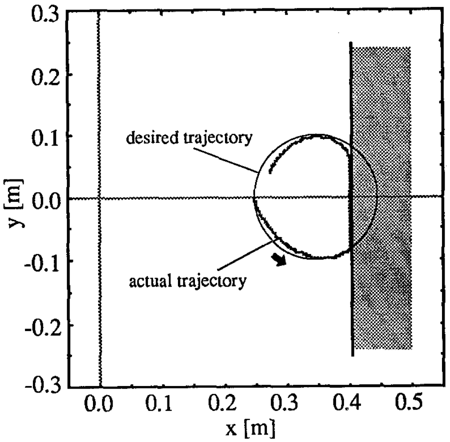


Fig. 11 Trajectory of free movement & Contact task **after learning**

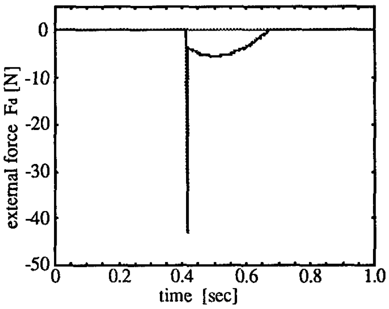


Fig. 12 External force at the end effector of free movement & Contact task **after learning**

4. Animal adaptive motor control

Human beings can stand stably upright because they observe the present state and control posture with a real-time feedback loop.

It has been pointed out that proprioceptors such as the muscle spindles and the tendon organs of Golgi, the vestibulo organ and vision all play a major part as sensors for observing the present state. The proprioceptors observe the position, velocity and force of the limbs and the body trunk. The vestibulo organs observe the head velocity and acceleration, and the visual sensor has the ability to sense the head velocity. Even if one sensor is seriously damaged, other sensors can compensate in order to maintain upright posture control. It is known that there are some descending tracts in the spinal cord from the brain-stem and the cerebellum to limb and body trunk muscles to harmoniously control posture and locomotion.[13]

On the other hand, Ito [14] has revealed physiologically that the cerebellum plays an important role in animal motor control adaptation and learning. There are four main parts of the cerebellum: the flocculus, the vermis, the hemisphere intermediate part and the hemisphere lateral part. Fujita [15] proposed the hypothesis of adaptive mechanism of vestibulo-ocular reflex in the flocculus, and Kawato [6] proposed the hypothesis of adaptive mechanism of voluntary movement in the hemisphere lateral part. Nashner [13] found that the adaptation of posture control is severely impaired in patients with cerebellar disease. Thus, we can imagine that the cerebellum plays an important role in the adaptation of posture, too.

The inputs and outputs of these four parts have already been investigated in anatomical studies.[16] The principal afferent inputs of the vermis in the cerebellum are from the vestibular labyrinth, the proximal body parts and the visual organs. The principal outputs of the vermis are to the medial brain stem and the axial regions of the motor cortex. The function of the vermis is comprehended as axial and proximal motor control. The hemisphere intermediate part gets the afferent information from the spinal cord, and sends the outputs to the red nucleus and the distal region of the motor cortex. The function of the hemisphere intermediate part is comprehended as distal motor control.

These facts support the hypothesis that the vermis and the hemisphere intermediate part have an adaptive mechanism of posture control and locomotion control. In addition to the previous two models by Fujita and Kawato, we propose two other models of adaptive control in the cerebellum for posture and locomotion control, as shown in Fig.13. These adaptive mechanisms are based on *feedback-error-learning* technique. Two adaptive models of the vermis and the hemisphere intermediate part correspond to the learning scheme that we proposed above.

5.Conclusion

We proposed here a new learning scheme using *feedback-error-learning* for a neural network model applied to adaptive nonlinear feedback control. After the neural network compensated perfectly or partially for the nonlinearity of the controlled object through learning, the response of the controlled object finally follows the desired response set in the conventional feedback controller. This learning scheme does not require the knowledge of the nonlinearity of a controlled object in advance. We can, therefore, use this learning scheme for many kinds of controlled objects, such as chemical plants, machines, robots, etc., because it is a very general learning scheme using nonlinear compensation technique.

Acknowledgment

We would like to thank Drs. E. Yodogawa and K. Nakane of ATR Auditory and Visual Perception Research Laboratories for their continuous encouragement.

References

[1] Dubowsky, S., DesForges,D. T. : The Application of Model Referenced Adaptive Control to Robotic Manipulators; Journal of Dynamic Sys., *Meas. and Cont.*Vol.101, pp.193-200 (1979)

[2]Slotine, J.E., Li, W.: On the adaptive control of robot manipulators, *Int. J. Robotics Research,* 6(3) (1987)

[3]Barto, A.G., Sutton, R.S., Anderson, C.W.: Neuronlike adaptive elements that can solve difficult learning control problems; *IEEE Trans. on Sys.,Man, and Cyb.* SMC-13, pp.834-846 (1983)

[4]Jordan, M.I.: Supervised learning and systems with excess degrees of freedom. COINS Technical Report 88-27, pp.1-41 (1988)

[5]Psaltis, D., Sideris, A., Yamamura, A.: Neural Controllers, IEEE Int. Conf. Neural Networks, Vol.4, pp.551-557 (1987)

[6]Kawato, M., Furukawa, K., Suzuki, R.: A Hierarchical Neural-Network Model for Control and Learning of Voluntary Movement; Biol. Cybern. 57, pp.169-185 (1987)

[7]Hogan, N.: Impedance Control: An Approach to Manipulation: Part I - Theory, PartII - Implementation, Part III - Applications, ASME Journal of Dynamic Systems, Measurement, and Control, Vol.107, pp.1-24 (1985)

[8]Kawato, M.: The feedback-error-learning neural network for supervised motor learning. In: Eckmiller, R.(ed) *Neural Network for Sensory and Motor Systems.* Elsevier, Amsterdam, pp.356-372 (1990)

[9]Rumelhart, D. E., McClelland, J.L. and PDP Research Group: Parallel Distributed Processing Vol.1, The MIT Press (1986)

[10]Albus, J.S.: A new approach to manipulator control: The cerebellar model articulation controller, Trans. ASME, vol.97 220 (1975)

[11]Stanfill, C., Waltz, D.: Toward memory-based reasoning, Communications of the ACM, vol.29, pp.1213-1228 (1986)

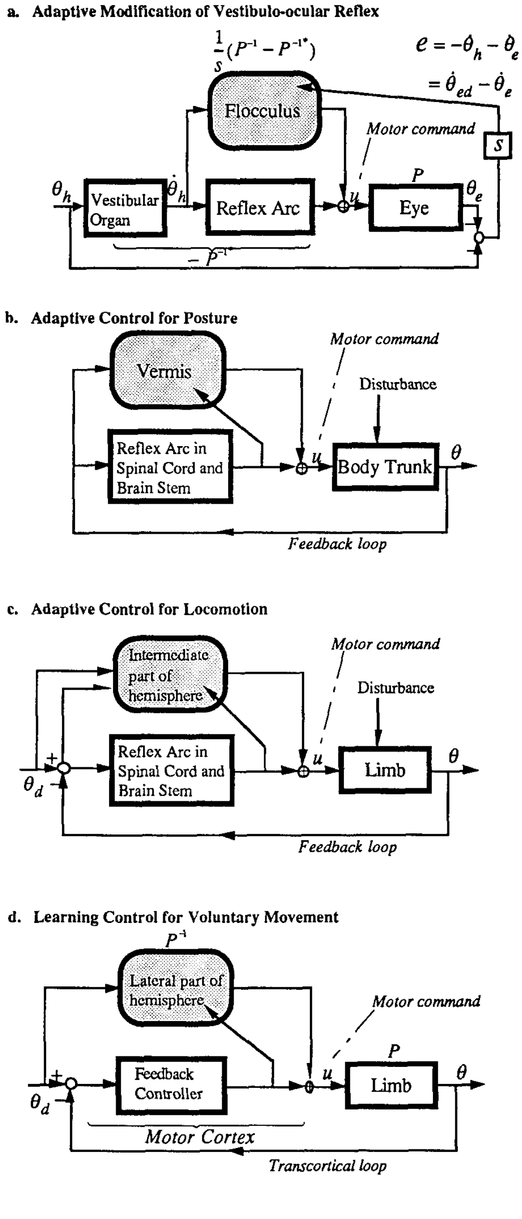
[12]Geman, S.: Some averaging and stability results for random differential equations; SIAM J. Appl. Math. vol.36 No.1, pp.86-105 (1979)

[13]Carew, T.J.:Posture and Locomotion. In : Kandel, E.R.(ed) Principles of Neural Science, Section 37, Elsevier Inc., 1985.

[14]Ito, M.: *The Cerebellum and Neural Control, Raven Press,* Section 25, pp.354-373 (1984)

[15]Fujita, M. : Simulation of adaptive modification of the vestibulo-ocular reflex with an adaptive filter model of the cerebellum, *Biol. Cybern.*, vol.45, pp.207-214 (1982)

[16]Ghez, C., Fahn, S. : The Cerebellum. In : Kandel, E.R.(ed) *Principles of Neural Science,* Section 39, Elsevier Inc.(1985).



a. Adaptive Modification of Vestibulo-ocular Reflex

b. Adaptive Control for Posture

c. Adaptive Control for Locomotion

Intermediate part of hemisphere
Motor command
Disturbance
+
θ_d
Reflex Arc in Spinal Cord and Brain Stem
u
Limb
θ
Feedback loop

d. Learning Control for Voluntary Movement

Fig. 13 Adaptive motor control in cerebellum