Biol. Cybern. 45, 195–206 (1982)

Biological Cybernetics

Adaptive Filter Model of the Cerebellum

M. Fujita

Department of Administration Technology, Nagasaki Institute of Applied Science, Nagasaki-shi, Nagasaki, Japan

Abstract. The Marr-Albus model of the cerebellum has been reformulated with linear system analysis. This adaptive linear filter model of the cerebellum performs a filtering action of a phase lead-lag compensator with learning capability, and will give an account for the phenomena which have been termed "cerebellar compensation". It is postulated that a Golgi cell may act as a phase lag element; for example, as a leaky integrator with time constant about several seconds. Under this assumption, a mossy fiber – granule cell – Golgi cell input network functions as a phase lead-lag compensator. Output signals from Golgi-granule cell systems, namely, parallel fiber signals, are gathered together through variable synaptic connections to form a Purkinje cell output. From a general theory of adaptive linear filters, learning principles for these modifiable connections are derived. By these learning principles, a Purkinje cell output converges to the "desired response" to minimize the mean square error of the performance. In a more general sense, a Purkinje cell acquires a filtering function on the basis of multiple pairs of input signals and corresponding desired output signals. The mode of convergence of the output signal is described when the input signal is sinusoidal.

1. Introduction

During the 1960's the neuronal circuitry of the cerebellum was analyzed in great detail and the cerebellum was conceived to have computer-like functions, particularly as a smooth and effective controller of movement (Eccles et al., 1967). In response to accumulated experimental data, a number of constructive models of the cerebellum have been proposed. Marr (1969) and Albus (1971) proposed spatial pattern-classifier models with learning capabilities, based upon distinct roles of two different types of afferents to the cortex. However, these models do not account satisfactorily for processing of time-analog signals conveyed by frequency-modulated nerve impulses which carry information in the central nervous system. Calvert and Meno (1972) postulated a filter model which enhances the content of input signals at higher frequency range both in time and space. Hassul and Daniels (1977) proposed a model in which the cerebellar cortex acts as a phase lead or lag compensator. These latter models can deal with time-analog signals conveyed by nerve impulses, but they do not account for cerebellar learning capabilities, which are known in terms of cerebellar compensation (Ito, 1975).

This article proposes a new model which accounts for both learning and filtering actions of the cerebellum. The model is based on the theory of adaptive filters (Sakrison, 1963; Widrow et al., 1967, 1975; Tsypkin, 1971). It is assumed that a Golgi cell may act as a phase lag element; for example, as a leaky integrator with time constant on the order of several seconds. By this assumption, output signals from granule cells in a mossy fiber – granule cell – Golgi cell input network are equivalent to the outputs from different versions of phase lead-lag compensators. These signals, namely, parallel fiber signals, are gathered together through adjustable synaptic weights to form a final output signal, a Purkinje cell output. The Marr-Albus type of learning principle is introduced in combination with the adaptive linear filter theory. Learning processes of the model are analyzed on the basis of the principle of stochastic approximation, which has already been applied to various learning schemes such as associative memory (Amari, 1977), formation of feature extracting cells (Amari and Takeuchi, 1978) and topographic organization of nerve fields (Amari, 1980).

0340-1200/82/0045/0195/$02.40

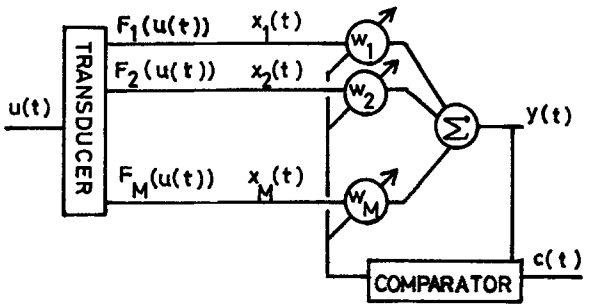

Fig. 1. Diagram of an adaptive linear filter

2. Cerebellar Cortex as Adaptive Linear Filter

2.1. Adaptive Linear Filter

A signal flow in an adaptive linear filter consists of two steps (Fig. 1). First, some transducer acts on an input signal $u(t)$ at time t to produce a linearly independent set of signals. Let $x_1(t), x_2(t), \ldots, x_M(t)$ be these intermediate signals at time t. They are gathered through adjustable weights w_j's and summed up to form a final output signal $y(t)$ at time t. That is,

$$x_j(t) = F_j(u(t)) \tag{2.1}$$

and

$$y(t) = \sum_{j=1}^{M} w_j x_j(t), \tag{2.2}$$

where F_j denotes the function of a transducer to the j-th output signal. With fixed values of these weights, the system outputs a filtered signal of an input. The term "adaptive" means that there exists a mechanism by which the w_j can be adjusted iteratively so that the filter output will approximate a given desired signal $c(t)$ as well as possible.

The mechanism is essentially equivalent to "forced learning" in the theory of neural models (Amari, 1977). That is, when the error signal $z(t)$ is given by

$$z(t) = y(t) - c(t), \tag{2.3}$$

learning processes on w_j can be described as

$$dw_j(t) = -\varepsilon z(t) x_j(t) dt, \tag{2.4}$$

where $dw_j(t)$ is the amount of weight change in $w_j(t)$ at time t and ε is a sufficiently small positive constant. We will consider the situation where an input signal $u(t)$ accompanied by the desired output signal $c(t)$ in time-duration T appears repetitively. Let the amount of weight change in w_j after each iteration be denoted by Δw_j, then from (2.4) it follows that

$$\Delta w_j = -\varepsilon \int_0^T z(t) x_j(t) dt. \tag{2.5}$$

That is, if the correlation between $z(t)$ and $x_j(t)$ is positive [i.e., $x_j(t)$ contributes to the error], then w_j through which $x_j(t)$ passes will be diminished. Otherwise, if the correlation is negative, then w_j will be increased. Characteristics of this learning principle will be discussed in Sect. 3, which concludes that adaptive changes in w_j's will progress to minimize the mean square error of the output signal. Especially if the desired signal $c(t)$ is given by

$$c(t) = \sum_{j=1}^{M} c_j x_j(t), \tag{2.6}$$

(i.e., a desired output signal is realizable with proper values of w_j's), then the mean square error will converge to 0. Furthermore, when an input signal accompanying a desired output signal appears randomly, the mean square error of the performance will also be minimized through the above learning processes.

This transducer corresponds to Golgi-granule cell systems, and the final output unit and the error signal correspond to a Purkinje cell and a climbing fiber signal, respectively, in the present model of the cerebellum.

2.2. Golgi – Granule Cell System

All physiological and anatomical data without references to sources are taken from Eccles et al. (1967) and Palkovits et al. (1971 A, B). The cerebellar cortex consists of five types of neurons, namely, Purkinje, basket, superficial stellate, Golgi, and granule cells. It receives two types of distinctively different afferents, mossy fibers and climbing fibers, and it sends only one type of efferent, Purkinje cell axons, to cerebellar and vestibular nuclei. There are about 330 Purkinje cells per mm^2 in the cortex (Palkovits et al., 1971 A). Each Golgi cell occupies $0.002\,mm^3$ of the granular layer where about 5700 granule cells and 3 Purkinje cells are present (Palkovits et al., 1971 B). The average granule cell has 3–5 dendrites; The range is 1–7. The axon of a granule cell ascends to the molecular layer of the cortex, where it bifurcates to form parallel fibers, which run along the folium for about 3 mm. A Purkinje cell has a fan-like dendritic arborization that is confined strictly to a plane perpendicular to the long axis of a folium. Thus, parallel fibers are orthogonal to the dendrites of Purkinje cells, and they make crossing-over synapses.

A Golgi cell receives signals both from mossy fibers and parallel fibers, and in turn inhibits granule cells (Fig. 2A). Since it can be assumed that dendritic expansions and axonal ramifications of neighbouring Golgi cells have neither any major overlaps nor large gaps, i.e., each Golgi cell occupies a tissue compartment (Eccles et al., 1967), each Golgi-granule cell

system may function as a unit of the input system. Morphological considerations about the density of Golgi cells (Eccles et al., 1967; Palkovits et al., 1971; Lange, 1974) suggest that a number (between a few tens and several hundreds) of Golgi-granule cell systems send parallel fiber signals to dendrites of a Purkinje cell.

A schematic diagram of a Golgi-granule cell system is shown in Fig. 2B. In the analysis of this neuronal network with linear system theory, neuronal signals are represented by discharge rates (impulses/s). Variables in small letters denote the deviational part of discharge rates from the spontaneous level and capital letters denote the total discharge rates. Output of the Golgi cell, v, and that of the j-th granule cell, x_j, each representing the variational part of discharge rates, are respectively derived on the basis of the diagram in Fig. 2B, such that

$$v=G\left(\sum_{k=1}^{L} b_k u_k+\sum_{k=1}^{L} q_k x_k\right) \tag{2.7}$$

and

$$x_j=g_j(a_j u_j-d_j v), \tag{2.8}$$

where a time function and the corresponding Laplace transform are represented by the same symbol for simplicity.

The following assumptions are necessary for further analyses.

[A1] A Golgi cell functions as a phase lag element. Especially if it functions as a leaky integrator with time constant T, its transfer function will be expressed simply as

$$G(s)=K/(1+Ts), \tag{2.9}$$

where $K>0$ and s denotes the Laplace transform of complex frequency (Discussion).

[A2] All the granule cells function as proportional elements with an identical gain value $g>0$ (Discussion).

[A3] Mossy fiber signals u_j's to a Golgi-granule cell system are of one type, i.e., $u_j=u$ for all values of j (Discussion).

[A4] A Golgi cell never receives parallel fiber signals from other Golgi-granule cell systems, as in Fig. 2B, which avoids further complications in the analysis.

From [A2] and [A3], it turns out that

$$x_j=g(a_j-d_jH)u \tag{2.10}$$

and

$$H=\left(g\sum_{k=1}^{L} q_k a_k+\sum_{k=1}^{L} b_k\right)G\Big/\left[1+\left(g\sum_{k=1}^{L} q_k d_k\right)G\right]. \tag{2.11}$$

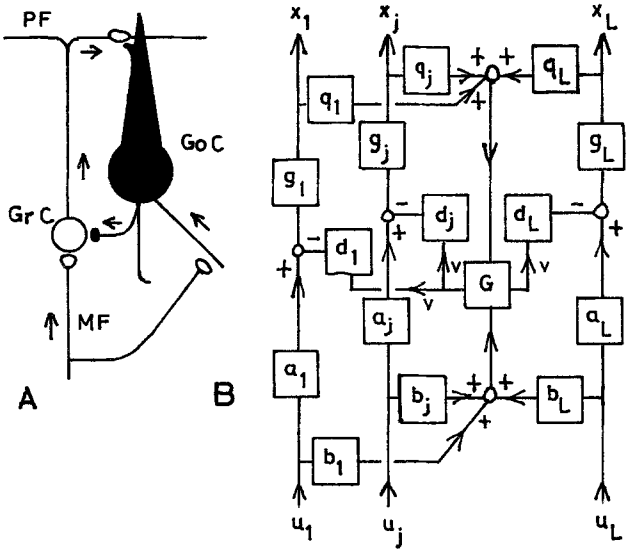

Fig. 2A and B. Schematic diagram of a neuronal circuit for a Golgi cell. **A** "The mossy fiber (MF) is input to the negative feedback circuit via granule cells (Gr C), and so return inhibition to granule cells. Also there is the mossy fiber synapse directly on the Golgi cell (Go C)" (from Fig. 119B in Eccles et al., 1967). **B** Block diagram of a Golgi-granule cell system. G and g_j ($j=1, 2, \ldots, L$) represent transfer functions of a Golgi cell and the j-th granule cell attached to it, respectively. u_j and x_j express the j-th mossy fiber input and the j-th parallel fiber output respectively. Connection weights between these cells and fibers are designated by a_j, b_j, q_j, and d_j

Although (2.10) is almost equivalent to what Albus (1971) gave in analyzing the same system, it shows quite different properties of the system from linear system theoretical viewpoint. The gain elements a_k, b_k, q_k, and d_k can be assumed ordinarily to take positive values. Equation (2.11) indicates that H is a transfer function of a phase lag element so long as G is that of a phase lag one. Therefore x_j is a phase lead version of the input signal u.

The dynamics of the system will be more easily understood by substituting (2.9) into (2.11), and for (2.10)

$$x_j=a_j g\left[1+r_j Kg\left(\sum_{k=1}^{L} q_k d_k\right)+Ts\right] \cdot u\Big/\left[1+Kg\left(\sum_{k=1}^{L} q_k d_k\right)+Ts\right], \tag{2.12}$$

where

$$r_j=1-\left[d_j\sum_{k=1}^{L}(b_k+g a_k q_k)\right]\Big/\left[a_j\sum_{k=1}^{L}(g q_k d_k)\right]. \tag{2.13}$$

Equation (2.12) means that this system functions as the most simple phase lead-lag compensator in a control system. When $r_j<1$, the system is a phase lead compensator, and when $1<r_j$, then a phase lag compensator (Discussion).

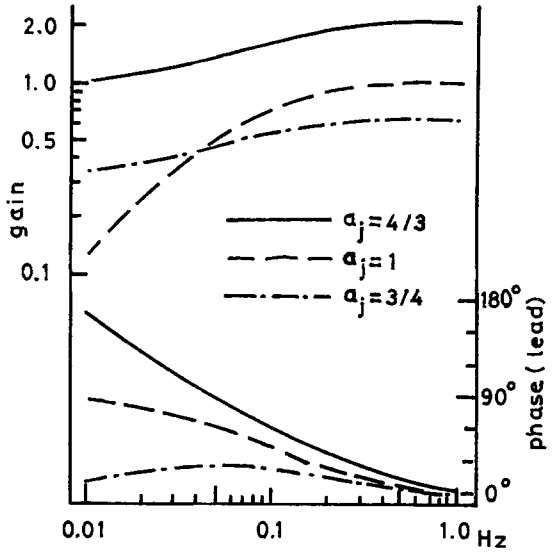

Fig. 3. Bode diagrams of output signals from a transducer, a Golgi-granule cell system, postulated in the cerebellar circuitry. Lines are plotted for the case when $T=4.0$(s) in (2.16)

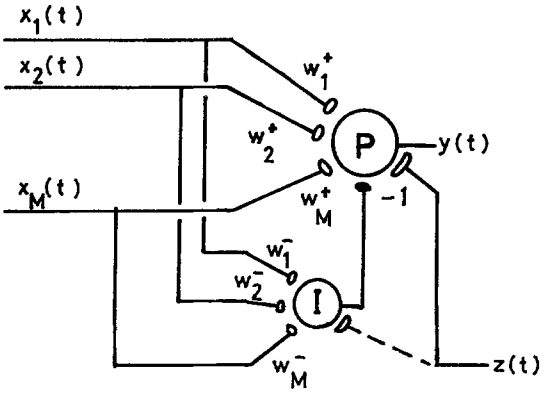

Fig. 4. Schematic diagram of a neuronal circuit for a Purkinje cell. *P* and *I* denote a Purkinje cell and an inhibitory interneuron model respectively, and a connection weight between them is assumed to take -1. A climbing fiber signal $z(t)$ is also provided to a Purkinje cell which functions as a "teacher signal" to adjust a modifiable weight w_j^+ $(j=1, 2, \ldots, M)$. Since it is assumed in the model that $z(t)$ does not modify connection weight w_j^- $(j=1, 2, \ldots, M)$, this collateral pathway is designated by a broken line

Furthermore, if parametric values in (2.13) such as a_j vary for different values of j, then the signals x_j's $(j=1, \ldots, L)$ will be those from different version of phase lead-lag compensators. This is what is necessary for the first step of signal flow in an adaptive linear filter. As b_k and q_k in (2.12) and (2.13) appear only in the total sum, it may well be assumed that $b_k=b$ and $q_k=q$ for all values of k. Then

$$r_j=1-(d_j/a_j)(b+g\bar{a}q)/(gq\bar{d}), \tag{2.14}$$

where $\bar{a}=(\sum a_k)/L$ and $\bar{d}=(\sum d_k)/L$.

While variation of d_j seems to have no basis for discussion, variability in a_j is plausible for the following reason. A nonlinear property of granule cells which may relate to the idea of a "codon cell" (Marr, 1969) is that summation of the synaptic excitations on at least two of four or five claws is required to evoke a granule cell discharge (Eccles et al., 1967). From a linear systems theoretical viewpoint, it might be predictable that the frequency modulation of discharge rates of a granule cell, synchronous to mossy fiber input signals, will be enhanced as the number of arrivals of input signals to its claws increases. This means in the model (Fig. 2B) that the transmission efficacy a_j may vary for different values of j.

To acquire quantitative feeling about the parameters in the system, one possible and very simple estimation is given in the Appendix 1. This shows that

$$v=2u/(2+Ts) \tag{2.15}$$

and

$$x_j=(a_j-2/(2+Ts))u, \tag{2.16}$$

where T takes value from 2.5 to 13.0(s). The Bode diagrams of x_j's when $a_j=4/3$, 1.0 or 3/4 (Fig. 3) show critical differences in phase among the low frequency responses of x_j's with a_j around 1.0.

Thus, Golgi-granule cell systems constitute a transducer of an adaptive linear filter. When a sinusoidal signal

$$u(t)=\sin\phi t \tag{2.17}$$

is input to the system with angular frequency ϕ, then both the phases θ_j's and gains ϱ_j's of the x_j's will vary:

$$x_j=\varrho_j\sin(\phi t+\theta_j). \tag{2.18}$$

Experiment results suggest indirectly that there is phase diversification in parallel fiber signals (Discussion).

2.3. Inhibitory Cells for Purkinje Cells

Output signals from Golgi-granule cell systems reach Purkinje cells directly, or indirectly via inhibitory interneurons, basket and superficial stellate cells. This part of neural network model is shown in Fig. 4, where complex pathways to Purkinje cells are simplified by assuming one inhibitory interneuron for one Purkinje cell with the aid of the law of superposition in a linear system. The connection weight of an intermediate signal $x_j(t)$ to this inhibitory interneuron model is denoted by w_j^-, and the weight of direct synaptic contact to a Purkinje cell by w_j^+. Then the final output $y(t)$ will be expressed by

$$\begin{aligned} y(t) &= \sum_{j=1}^{M} w_j^+ x_j(t) - \sum_{j=1}^{M} w_j^- x_j(t) \\ &= \sum_{j=1}^{M} w_j x_j(t), \end{aligned} \tag{2.19}$$

where $w_j = w_j^+ - w_j^-$ for $j=1, \ldots, M$ and M is the number of signals which reach the Purkinje cell from Golgi-granule cell systems.

2.4. Climbing Fibers and Learning Principle

A climbing fiber, originating from the inferior olive, exerts a powerful excitatory action on one Purkinje cell and induces characteristic burst discharges called complex spikes. Climbing fiber responses of a Purkinje cell occur much less frequently than the mossy fiber responses, represented by simple spikes. The spontaneous rate of the complex spikes is 1–2 impulses/s (Ghelarducci et al., 1975). It seems likely that the mossy fiber responses represent the output of information processing in the cerebellum, and the climbing fiber responses play some other role, which Marr (1969) and Albus (1971) assumed as that of a "teacher signal".

Let $Z(t)$ be the activity of a climbing fiber at time t, and put

$$Z(t) = z_0 + z(t), \tag{2.20}$$

where z_0 denotes the spontaneous discharge rate and takes a constant value. It is now assumed that $z(t)$, a deviational part of the climbing fiber impulse rate, is given by the difference between an output signal $y(t)$ and a desired response $c(t)$; $z(t)$ is defined by (2.3). In other words, $c(t)$ can be defined implicitly by (2.3). Since (2.19) is equivalent to (2.2) in an adaptive linear filter, the learning principle (2.4) is also applicable to the present case. The present model can not identify changes in w_j with those in w_j^+ and/or in w_j^-, and a variety of learning principles are possible (Discussion).

The following arguments are under the assumption that $dw_j^-/dt = 0$, that is, w_j^- takes large enough constant value so that changes in w_j^+ can produce necessary modification of w_j. After (2.4), it follows directly that

$$dw_j^+(t) = -\varepsilon z(t) x_j(t) dt, \quad (\varepsilon > 0), \tag{2.21}$$

$$dw_j^-(t)/dt = 0. \tag{2.22}$$

Since the heterosynaptic interaction hypothesis (Marr, 1969; Albus, 1971; Ito, 1972; Eccles, 1977) suggests that it might be unreasonable that identical phenomena would occur in both cases when $z(t) > 0$, $x_j(t) > 0$ and when $z(t) < 0$, $x_j(t) < 0$, a modified version of (2.21) will be adopted for learning principle as follows:

$$dw_j^+(t) = -\varepsilon z(t)[x_j(t) + x_0] dt, \quad (\varepsilon > 0), \tag{2.23}$$

where $dw_j^+(t)$ is the amount of weight change in $w_j^+(t)$ at time t, and x_0 denotes a spontaneous discharge rate.

Learning Principle. A synaptic weight of the j-th parallel fiber to a Purkinje dendrite decreases when $Z(t)$ is greater than z_0, and increases when $Z(t)$ is smaller than z_0. The amount of the weight change is proportional to the product of the j-th parallel fiber impulse rate and $z(t)$, the deviational part of $Z(t)$ from z_0.

In spite of the discrepancy in the term x_0 between (2.21) and (2.23), they may be equivalent in the real neuronal system, possibly by adjusting the activity level of neuronal discharges, such that

$$\int_0^T z(t) dt = 0. \tag{2.24}$$

Equation (2.4) or (2.21) is assumed in the following development of the model.

3. Characteristics of Adaptive Learning

3.1. Learning for a Single Input Signal

First, we will consider the case of the repetitive arrival of a single signal $u(t)$, accompanied by the desired response $c(t)$ over a time-duration T. Each time, the signal $u(t)$ will be converted to $x_j(t)$'s and the effects of learning processes on w_j will follow with the help of $z(t)$. Let $w_j[n]$ be the value of w_j after n times repetition of the learning process. Then it follows from (2.5) that

$$w_j[n+1] - w_j[n] = -\varepsilon \int_0^T \left(\sum_{k=1}^M w_k(t) x_k(t) - c(t) \right) x_j(t) dt. \tag{3.1}$$

Furthermore, it is assumed that ε is so small that the changes in $w_j(t)$, and hence in $y(t)$, are negligibly small in each iteration. This allows the replacement of $w_k(t)$ in (3.1) by $w_k[n]$ as

$$w_j[n+1] - w_j[n] = -\varepsilon \int_0^T \left(\sum_{k=1}^M w_k[n] x_k(t) - c(t) \right) x_j(t) dt. \tag{3.2}$$

The limiting values of w_j in (3.1) and (3.2) are clearly equal, when they exist. With vector notation, (3.2) can be expressed as

$$\mathbf{w}[n+1] = \mathbf{w}[n] - \varepsilon(R\mathbf{w}[n] - \mathbf{b}), \tag{3.3}$$

where

$$\mathbf{w} = (w_1, w_2, \ldots, w_M)^*, \tag{3.4}$$

$$\mathbf{b} = \int_0^T c(t) \mathbf{x}(t) dt, \tag{3.5}$$

$$\mathbf{x}(t) = (x_1(t), x_2(t), \ldots, x_M(t))^* \tag{3.6}$$

are column vectors,

$$R = \int_0^T \mathbf{x}(t) \mathbf{x}^*(t) dt \tag{3.7}$$

is an $M \times M$ matrix, and "*" denotes the transposition of a vector.

Let the criterion of the system performance be the mean square error, which is expressed by a function of weights w_j's, that is,

$$J(w_1, w_2, \ldots, w_M) = (1/2T) \int_0^T (y(t) - c(t))^2 dt . \tag{3.8}$$

By partial differentiation with respect to w_j, it holds that

$$\partial J / \partial w_j = (1/T) \int_0^T (y(t) - c(t)) x_j(t) dt . \tag{3.9}$$

Comparison with (3.2) gives

$$w_j[n+1] - w_j[n] = -\varepsilon T (\partial J / \partial w_j) . \tag{3.10}$$

This means that during the adaptive process of w_j, if $\varepsilon > 0$, then J declines to its minimum by descending in the steepest direction, i.e., the modification of w_j follows the steepest descent method (e.g. Tsypkin, 1971). The function J has a single minimum, because J is a quadratic form in w_j with a positive-definite or positive semi-definite symmetric coefficient matrix R.

3.2. *Learning for an Ensemble of Input Signals*

Next, a random information source Ω will be considered, which includes input signals $u(t, \omega)$ with time-duration T and the associated desired signals $c(t, \omega)$, where $\omega \in \Omega$ is a parameter specifying the signals. In this case, the mean square error in (3.8) will be expressed as

$$J(w_j, \omega) = (1/2T) \int_0^T z(t, \omega)^2 dt , \tag{3.11}$$

where

$$z(t, \omega) = \sum_{k=1}^{M} w_k x_k(t, \omega) - c(t, \omega) . \tag{3.12}$$

A function $J(w_j, \omega)$, a multi-dimensional parabolic hyperplane of w_j's, changes on ω. Thus at each step of adaptation, the descending direction given by (3.10) depends not only on w_j, but also on the signal $u(t, \omega)$ accompanied by the signal $c(t, \omega)$ from the source, in other words, on what parabolic hyperplane $J(w_j, \omega)$ is selected. However, provided that input signals $u(t, \omega)$ are produced from the stationary source independently at each time with some probability on ω, w_j will converge to some unique optimal value. If ε is a sufficiently small positive constant, the descending route of w_j will coincide on the average, after many times repetition, with the steepest descent route of the expected function $E[J(w_j, \omega)]$, where $E[\cdot]$ denotes the expectation with respect to $\omega \in \Omega$. This is well-known as stochastic approximation (Wasan, 1969).

The simple convergence properties of the learning process by (2.4) or (3.2) will be presented next. Under the conditions:

i) A source Ω contains a finite number of signals $u(t, \omega)$ and $c(t, \omega)$,

ii) At each time, an input signal $u(t, \omega)$ with the associated desired signal $c(t, \omega)$ is issued independently from the source Ω,

iii) For each ω, M signals $\{x_j(t, \omega)\}$ $(j = 1, \ldots, M)$ are linearly independent,

iv) $\varepsilon > 0$ is sufficiently small,

then the weight vector $\mathbf{w}$ converges, irrespective of its initial value, to $\mathbf{w}_{\text{opt}}$ in the sense that

$$\lim_{n \to \infty} E[\mathbf{w}[n]] = \mathbf{w}_{\text{opt}}, \tag{3.13}$$

$$\lim_{n \to \infty} E[\|\mathbf{w}[n] - \mathbf{w}_{\text{opt}}\|^2] = O(\varepsilon), \tag{3.14}$$

where $O(\varepsilon)$ is the term of order ε and $\mathbf{w}_{\text{opt}}$ is the optimum vector which minimizes

$$E[J(\mathbf{w}, \omega)] = \mathbf{w}^* E[R(\omega)] \mathbf{w} - 2\mathbf{w}^* E[\mathbf{b}(\omega)] + \int_0^T c(t)^2 dt . \tag{3.15}$$

Furthermore, if, instead of the condition (ii), a desired signal $c(t, \omega)$ can be expressed by a given vector

$$\mathbf{c} = (c_1, c_2, \ldots, c_M)^* \tag{3.16}$$

as

$$c(t, \omega) = \sum_{k=1}^{M} c_k x_k(t, \omega), \tag{3.17}$$

[i.e., the desired response signal is obtained by filtering the input signal $u(t, \omega)$], then it follows that

$$\lim_{n \to \infty} E[\mathbf{w}[n]] = \mathbf{c}, \tag{3.18}$$

$$\lim_{n \to \infty} E[\|\mathbf{w}[n] - \mathbf{c}\|^2] = 0 . \tag{3.19}$$

Note that in this case

$$E[J(\mathbf{w}, \omega)] = (\mathbf{w} - \mathbf{c})^* E[R(\omega)] (\mathbf{w} - \mathbf{c}) . \tag{3.20}$$

When condition (iii) is not satisfied, the optimum vector $\mathbf{w}_{\text{opt}}$, which still minimizes $E[J(\mathbf{w}, \omega)]$, will not be unique but will depend on its initial value. Outlines of proofs of these properties are given in Appendix 2.

3.3. *Mode of Convergence in Adaptive Learning*

The convergence of the system performance under the learning principle is demonstrated in the preceding sections. If the original input signal is sinusoidal, then

the mode of convergence of the output signal to a desired signal can be formulated. Since a sinusoidal signal in a linear system can be denoted by the amplitude and phase, a polar representation of a signal is employed for convenience. Let $u(t)$ be expressed by

$$u(t)=\exp(i\phi t), \tag{3.21}$$

where $i^2=-1$ and ϕ denotes the angular frequency of an input signal $u(t)$. Then from (2.12) the k-th output of a transducer $x_k(t)$ will be expressed with constants ϱ_k and θ_k as

$$x_k(t)=\varrho_k\exp(i\theta_k)\exp(i\phi t), \tag{3.22}$$

and it follows that

$$\begin{aligned} y(t) &= \sum_{k=1}^{M} w_k x_k(t) \\ &= \sum_{k=1}^{M} w_k\varrho_k\exp(i\theta_k)\exp(i\phi t) \\ &= \Lambda\exp(i\Psi)\exp(i\phi t), \end{aligned} \tag{3.23}$$

with appropriate constants Λ and Ψ. In vector notation on the complex plane, (3.22) and (3.23) can be replaced by $\mathbf{x}_k$ and $\mathbf{y}$,

$$\mathbf{x}_k=\varrho_k\exp(i\theta_k) \tag{3.24}$$

and

$$\mathbf{y}=\Lambda\exp(i\Psi). \tag{3.25}$$

Let the desired response $c(t)$ be given as a filtered output of $u(t)$, as in (3.17),

$$c(t)=\sum_{k=1}^{M} c_k x_k(t). \tag{3.26}$$

With vector notation like in (3.23), it holds with some constants r and ψ that

$$\mathbf{c}=r\exp(i\psi). \tag{3.27}$$

The change in $y(t)$ corresponding to the change in w_k by (3.2) is presented as follows. Let the modified filter output by adaptive learning (3.2) be denoted by $\mathbf{y}+\Delta\mathbf{y}$, then the increment can be expressed as

$$\Delta\mathbf{y}=-\mu[(\mathbf{y}-\mathbf{c})+h(\mathbf{y}-\mathbf{c})^{\dagger}], \tag{3.28}$$

where "†" denotes the complex conjugate,

$$\mu=\varepsilon\left(\sum_{k=1}^{M}\varrho_k^2\right)^{T/4} \tag{3.29}$$

and

$$h=\left[\sum_{k=1}^{M}\varrho_k^2\exp(2i\theta_k)\right]\Big/\left(\sum_{k=1}^{M}\varrho_k^2\right). \tag{3.30}$$

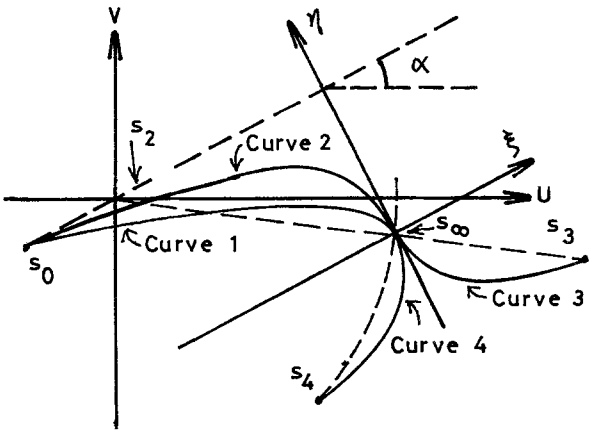

Fig. 5. Mode of convergence in adaptive learning in case of a sinusoidal input signal. The $U-V$ represents the Cartesian coordinate system of $\mathbf{y}$. The limiting value of $\mathbf{s}$ is denoted by $\mathbf{s}_\infty$, which equals to a desired signal $\mathbf{c}$ in this case. Curve *1*: from $\mathbf{s}_0$ to $\mathbf{s}_\infty$ with small β, Curve *2*: from $\mathbf{s}_0$ via $\mathbf{s}_2$ to $\mathbf{s}_\infty$ with large β, Curve *3*: from $\mathbf{s}_3$ to $\mathbf{s}_\infty$, Curve *4*: from $\mathbf{s}_4$ to $\mathbf{s}_\infty$, $\mathbf{s}_2$ is an intersection point of Curve *2* and a radius $\mathbf{Os}_\infty$

Furthermore, let h be expressed in terms of constants α and β as

$$h=\beta\exp(2i\alpha), \tag{3.31}$$

and set $\mathbf{s}$ as

$$\mathbf{s}=(\mathbf{y}-\mathbf{c})\exp(-i\alpha), \tag{3.32}$$

whose Cartesian coordinates are denoted by $\mathbf{s}=(\xi,\eta)$. Then the locus of the vector $\mathbf{s}$, and hence $\mathbf{y}$, is given by

$$(\eta/\eta_0)=(\xi/\xi_0)^{(1-\beta)/(1+\beta)}, \tag{3.33}$$

where $\mathbf{s}_0=(\xi_0,\eta_0)$ is the initial vector. Derivations of (3.28) and (3.33) are given in Appendix 3. From (3.30) and (3.31), α can be identified with the weighted mean value of the phase of intermediate signals. The effect of β is shown by the Curves *1* and *2* in Fig. 5. With small β, $\mathbf{y}$ is modified straightforwardly to $\mathbf{c}$, but with large β, the phase of $\mathbf{y}$ does not readily change. In the extreme example, the Curve *3* (or the Curve *4*) indicates inevitable phase (or gain) changes in $\mathbf{y}$ when $\mathbf{y}_0$ and $\mathbf{c}$ take identical phase (or gain) values but different gain (or phase) values.

Generally $0\leqq\beta\leqq 1$, and β is inversely related to the phase variance in intermediate signals. Let $\varrho_k=1$ for all values of k, and θ_k be uniformly distributed in an interval. Then, simple calculation shows that

$$\alpha=[\max(\theta_k)+\min(\theta_k)]/2, \tag{3.34}$$

$$\beta=\sin\Theta/\Theta, \tag{3.35}$$

where

$$\Theta(rad)=\max(\theta_k)-\min(\theta_k). \tag{3.36}$$

It is clear form (3.34) and (3.35) that α is the mean value of θ_k and β is inversely related to Θ.

4. Discussion

A phase lead or lag compensator is used to improve response characteristics of a feedback control system, by speeding up the response of a system, stabilizing a possibly unstable system, or reducing stationary errors of system performance (Takahashi et al., 1970). The present adaptive linear filter model of the cerebellum works as such a compensator, and it displays functional flexibility. These properties are consistent with data regarding cerebellar compensation from preparations which include the stretch reflex (Higgins et al., 1962), the pupillary light reflex (Tsukahara et al., 1973), and the vestibuloocular reflex (Ito, 1978). Although this model of the cerebellum is based upon a number of assumptions and simplifications, it can explain these simple reflexes under cerebellar control as more knowledge is obtained about the neuronal circuitry involved. The validity of these assumptions will be discussed in this section.

4.1. Linearity of the System

A neuron has some nonlinear properties in its impulse discharge rate, such as threshold and saturation. The present model is built on a linearity assumtion for signal processing in the cerebellum. Electrophysiological studies have revealed that sinusoidal head rotation induces sinusoidal frequency modulated impulses in vestibular nerve fibers, that are in-phase with head velocity (Fernandez and Goldberg, 1971). These impulses enter the cerebellar flocculus via mossy fiber afferents and induce sinusoidal frequency modulation of simple spike discharges of Purkinje cells (Lisberger and Fuchs, 1974, 1978; Ghelarducci et al., 1975). The sinusoidal wave of Purkinje cell output is cut off on the lower side only when the amplitude of modulation exceeds a limit. This observation ensures that the floccular circuitry operates as a linear system within a limited range of magnitude of input signals. Similar results were obtained by recording from vermal Purkinje cells during sinusoidal neck rotation (Denoth et al., 1980). Although these parts of the cortex are limited, the histologically uniform structure of the cerebellar cortex may suggest such linearity in the whole cortex.

4.2. The Golgi-Granule Cell System

4.2.1. Assumption of a Leaky Integrator. The fact that Golgi cell inhibition of granule cell discharges has a duration in excess of 100 ms (Eccles et al., 1967) and the observation of discharge patterns in a presumed Golgi cell (Miles et al., 1980) suggest that it serves as a phase lag element, mimicking the action of a leaky integrator. Future experiments will determine whether a Golgi cell usually shows a leaky integrator-like pattern with a time constant on the order of several seconds, and whether parallel fiber signals show frequency responses similar to those in Fig. 3.

4.2.2. Assumption of a Phase Lead. Since all parameters in (2.13) usually take positive values (Fig. 2A), then it holds that $r_j < 1$ and the system functions as a phase lead element. There is also the possibility that it is a phase lag element ($r_j > 1$). This would arise when either the numerator or denominator of the second term on right-hand side of (2.13) is negative. If the majority of a_k's and b_k's take negative values, then the numerator is negative. This is the case, for example, when a large number of mossy fibers bring signals to a flocculus from the contralateral canals, provided that these signals are out of phase with ipsilateral signals. Second, if the majority of q_k's are negative, i.e., strong signals are fed back to a Golgi cell via inhibitory interneurons from parallel fibers, but b_k's take sufficiently large positive values, then the denomenator is negative while the numerator is positive.

4.2.3. Gain Element Assumption for Granule Cells. There are no experimental data pertinent to this assumption. However, such property seems to be preferable for a relay cell, which receives excitatory effects of mossy fiber signals and inhibitory effects of Golgi cell outputs. The essential role of a granule cell seems to consist in its peculiar dendrites, as discussed in Sect. 2.

4.2.4. Identical Input Signals for a Golgi-Granule Cell System. This assumption may appear to be unnatural because mossy fibers from many sources converge onto each area of the cerebellar cortex. However, this assumption is not only for the sake of simplification of the model. It is possible that the minority of the input signals have the less chance to occupy the claws of granule cells so as to evoke discharges. These signals would be cut off by nonlinear properties of granule cells, while preponderant signals will be passed through. Even though input signals to one Golgi-granule cell system are constrained to be identical, the present model can be applied also to a case where several types of signals from different Golgi-granule cell systems may converge to a Purkinje cell, and that also the same learning principle as in (2.4) holds in this case from linearity.

4.2.5. Diversification in Phase of Parallel Fiber Signals. There have been no experimental observations which directly support this assumption. However, sinusoidal mossy fiber inputs (presumably with identical phase) evoke Purkinje cell discharges in various phase shifts, in the maximum more than several tens of degrees at 0.1 Hz (Ghelarducci et al., 1975; Dufosse et al., 1978; Denoth et al., 1980). The present adaptive filter model assumes that such phenomena are due to a variety of combinations of parallel fiber signals diversified in phase.

4.3. Learning Principles

The present model does not exclude the possibilities of a change restricted to w_j^-, or of changes in both w_j^+ and w_j^-, so long as (2.4) holds for the learning processes. The very recent findings by Ito et al. (1982) have shown that conjunctive stimulation of vestibular mossy fiber afferents at a rate of 20 Hz and climbing fiber afferents at 4 Hz causes a drastic depression of the responsiveness of flocculus Purkinje cells to vestibular mossy fiber afferents, just as Albus (1971) assumed. The present learning principle assumes both depression and reinforcement of the transmission efficay from parallel fibers to a Purkinje cell, which is partially consistent with the experiments. It may be noted that learning can be achieved similarly even when (a) w_j^+ and w_j^- both increase and decrease but with different values of time constants, or (b) w_j^+ or w_j^- changes in one direction (increase or decrease), with a natural restoration of the increased or decreased values of w_j^+ or w_j^- to standard levels. Equilibrium of these antagonistic forces realizes convergence of the learning process.

4.4. Microzone

Axons of superficial stellate type "a" neurons or basket cells extend transversely the folium, over about max. 0.9 mm or 1 mm in either direction, mediating the so-called off beam parallel fiber signals to a Purkinje cell. In addition, a climbing fiber, postulated in a model (Fig. 4) as an error signal cable, extends branches transversely across a folium (Eccles et al., 1967). Climbing fiber responses are correlated in time, and this correlation is extended perpendicular to the long axis of a folium (Bell and Kawasaki, 1972). Thus, it is desirable to develop a single Purkinje cell model in Fig. 4 to a model of Purkinje cells which cooperate and are aligned perpendicular to the long axis of a folium. In fact, recent topographical studies of the corticonuclear and olivocerebellar projection (Groenewegen and Voogd, 1977; Andersson and Oscarsson, 1978) led to the basic idea: A narrow longitudinal strip of cerebellar cortex (microzone) having a dimension of a few hundred micra by a few millimeters, serves as a functional unit of corticonuclear organization in the cerebellum. Even this small strip contains about half a million of neurons (including 500 Purkinje cells), and may represent one unit of a multipurpose computer (Ito, 1979).

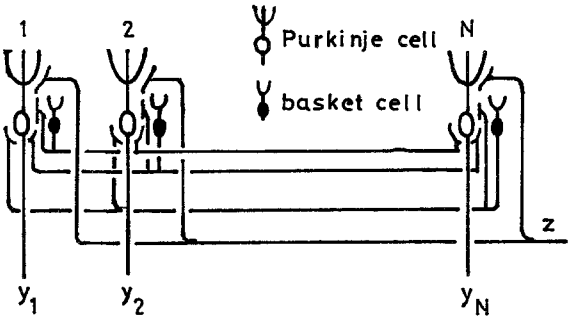

Fig. 6. A primitive model of microzone structure. N Purkinje cells and N basket cells are cooperating linearly, receiving identical climbing fiber signals. N output signals from Purkinje cells are summed up possibly in the nucleus. Parallel fibers are not shown for simplicity

Thus, the "one Purkinje cell model" may be expanded into a primitive model of a microzone (Fig. 6). The same learning capabilities hold by the law of superposition in a linear system. Although this model does not explain any more functioning of the cerebellum than a single cell model, it may be advantageous in the following sense.

(1) It is generally recognized that the characteristic dendrites of a Purkinje cell, arborizing orthogonal to the parallel fibers, improve the convergence number of input signals. In the same way, the model will receive variety of input signals. Microzones really extend perpendicular to the long axis of the folium as a model indicates.

(2) The abundant axonal distribution of inhibitory interneurons, superficial stellate cells and basket cells, may provide a microzone with inhibitory signals large enough to assume the selective adjustment of w_j^+ as in (2.21) and (2.22).

(3) In a microzone model it is easy to show that partial destruction of the Purkinje cells will be compensated by the remaining cells to maintain the total effect of output signals.

(4) Collaterals of cells of the deep cerebellar nuclei project to the cerebellar cortex as mossy fiber inputs (Ito, 1978). It is hopeful in future studies that the primitive microzone model will acquire much more flexible functioning, incorporating this type of mossy fiber input, which was neglected in a single cell model to avoid complexity.

Appendix 1. Estimation of the Parameters in (2.7)–(2.9)

The signals u, x_j, and v are variational parts of discharge rates, and hence it may not be unreasonable to assume their equal amplitudes in frequency responses. From (2.10) and (2.11), the gains of H and $g(a_j - dH)$ will be approximated by 1, and it will be estimated roughly that $g = d_k = 1$, $b_k = q_k = 0.5$ and $\sum a_k/L = 1$:

$$H(s) = KL/(1 + 0.5KL + Ts) \tag{A1-1}$$

and

$$x_j = [a_j - KL/(1 + 0.5KL + Ts)]u.$$

Set $\bar{x} = (\sum x_j)/L$, then

$$\bar{x} = (1 - 0.5KL + Ts)u/(1 + 0.5KL + Ts). \tag{A1-2}$$

The data of Miles et al. (1980) regarding observations of the signal v indicates by (A1-1),

$$T/(1 + 0.5KL) = 6.5(\text{s}).$$

A phase lead element was postulated in the cerebellar model by Gonshor and Melvill Jones (1976) to explain their experimental data. The present model suggests that the element is due to Golgi cell operation, and to satisfy their condition,

$$1 - 0.5KL = 0 \tag{A1-3}$$

and

$$T/(1 + 0.5KL) = 1.25(\text{s})$$

may hold from (A1-2). From (A1-3), $KL = 2$ follows and then T will take a value from 2.5 to 13.0 (s).

Appendix 2. Outlines of Proofs of Convergence Properties

The case where the condition (iii) does not hold is considered last in this section. Since the correlation vector and matrix corresponding respectively to (3.5) and (3.7) depend on ω in this case, let denote them by $\mathbf{b}(\omega)$ and $R(\omega)$. Let $\omega_n \in \Omega$ be a sample point for the n-th input signal. Then, similar to (3.2), it holds that

$$\mathbf{w}[n+1] = (I - \varepsilon R(\omega_{n+1}))\mathbf{w}[n] + \varepsilon \mathbf{b}(\omega_{n+1}), \tag{A2-1}$$

where I denotes the $M \times M$ unit matrix. By induction, it follows that

$$\mathbf{w}[n+1] = \left[\prod_{k=1}^{n+1} (I - \varepsilon R(\omega_k))\right]\mathbf{w}[0] + \sum_{k=1}^{n+1}\left[\prod_{j=k+1}^{n+1} (I - \varepsilon R(\omega_j))\right]\mathbf{b}(\omega_k), \tag{A2-2}$$

where the two terms on the right-hand side will be denoted by $\mathbf{u}[n+1]$ and $\mathbf{v}[n+1]$ respectively. Equations (3.13) and (3.14) can be derived by estimating the expectation

$$E[\mathbf{u}[n+1]]$$

and the euclidean norm

$$\|\mathbf{u}[n+1] - E[\mathbf{u}[n+1]]\|$$

as well as those of $\mathbf{v}[n+1]$ with the help of following remarks:

1) From the condition (ii), calculus of expected values will be carried out simply.

2) From the condition (i), $\|\mathbf{b}(\omega)\|$ is bounded for ω, i.e., there exists a positive number B, such that

$$\|\mathbf{b}(\omega)\| < B.$$

3) From the conditions (i) and (iii), the matrices $R(\omega)$ and $E[R(\omega)]$ are symmetric and positive definite, so that for any eigenvalue λ of these matrices there exist positive numbers $\lambda_{\min}$ and $\lambda_{\max}$, such that

$$0 < \lambda_{\min} < \lambda < \lambda_{\max}. \tag{A2-3}$$

Put $\bar{R}$ and $\bar{\mathbf{b}}$ as $\bar{R} = E[R(\omega)]$ and $\bar{\mathbf{b}} = E[\mathbf{b}(\omega)]$, then it can be shown easily that

$$\lim_{n \to \infty} (I - \varepsilon \bar{R})^{n+1} = 0,$$

$$\lim_{n \to \infty} E[\mathbf{w}[n+1]] = \bar{R}^{-1}\bar{\mathbf{b}},$$

and

$$\mathbf{w}_{\text{opt}} = \bar{R}^{-1}\bar{\mathbf{b}}.$$

Let ν be a positive number such that for sufficiently small ε

$$0 < \nu < 2\lambda_{\min} - \varepsilon\lambda_{\max}^2,$$

it holds that

$$E[\|\mathbf{w}[n+1] - E[\mathbf{w}[n+1]]\|^2] < 2\varepsilon B/\nu = O(\varepsilon).$$

Consider the case when $c(t, \omega) = \mathbf{c}^* \mathbf{x}(t, \omega)$ for

$$\mathbf{x}(t, \omega) = (x_1(t, \omega), x_2(t, \omega), \ldots, x_M(t, \omega))^*.$$

Since it holds in this case that

$$\mathbf{b}(\omega) = R(\omega)\mathbf{c},$$

substitution of the above into (A2-1) gives

$$\mathbf{w}[n+1] - \mathbf{c} = (I - \varepsilon R(\omega))(\mathbf{w}[n] - \mathbf{c}). \tag{A2-4}$$

From (A2-3) there exists γ $(0 < \gamma < 1)$, not depending on ω, such that

$$\|I - \varepsilon R(\omega)\| < 1 - \gamma < 1.$$

Then, taking the norms in both sides of (A2-4) shows the inequality (3.19).

In the case when the condition (iii) does not hold, assume that there exists a $M \times M$ matrix A independent of ω, such that

$$\mathbf{x}(t,\omega) = AP\mathbf{e}(t,\omega), \tag{A2-5}$$

where P is an arbitrary $m \times m$ regular matrix and

$$\mathbf{e}(t,\omega) = (e_1(t,\omega), e_2(t,\omega), \ldots, e_m(t,\omega))^*$$

is a m-dimensional column vector whose m components constitute a linearly independent basis of a vector space spanned by $x_j(t,\omega)$'s for $j=1,\ldots,M$. It will be shown by (A2-5) that

$$\begin{aligned} y(t,\omega) &= \mathbf{w}^*\mathbf{x}(t,\omega) \\ &= \mathbf{f}^*\mathbf{e}(t,\omega), \end{aligned}$$

where

$$\mathbf{f} = P^*A^*\mathbf{w}. \tag{A2-6}$$

Changes in $\mathbf{f}$ will be derived as in (A2-1), such that

$$\mathbf{f}[n+1] = (I - \varepsilon Q(\omega)D)\mathbf{f}[n] + \varepsilon \mathbf{d}(\omega)D,$$

where

$$Q(\omega) = \int_0^T \mathbf{e}(t,\omega)\mathbf{e}^*(t,\omega)dt,$$

$$\mathbf{d}(\omega) = \int_0^T c(t,\omega)\mathbf{e}(t,\omega)dt$$

and

$$D = P^*A^*AP.$$

Since A^*A is symmetric and positive definite, the matrix D becomes the $m \times m$ unit matrix when a proper regular matrix P is chosen. Changes in $\mathbf{f}$, thus, reduce to those when the condition (iii) holds. However, (A2–6) indicates that a vector $\mathbf{w}_{opt}$ can not be determined uniquely for the optimum vector $\mathbf{e}_{opt}$, due to the degeneracy in a parabolic hyperplane $E[J(\mathbf{w},\omega)]$ of w_i's. $\mathbf{w}_{opt}$ depends on the initial value.

Appendix 3. Derivations of (3.28) and (3.33)

Taking care to calculate the time correlation of signals in polar representation, integral calculus of (3.2) in terms of (3.22), (3.23), and (3.26) will be performed as follows.

$$\begin{aligned} \Delta w_k &= -\varepsilon \int_0^T [\Lambda \exp(i\Psi) \\ &\quad - r\exp(i\psi)]exp(i\phi t)0.5[\varrho_k \exp i(\theta_k + \phi t) \\ &\quad + \varrho_k \exp(-i(\theta_k + \phi t))]dt, \\ &= (-\varepsilon/2)\varrho_k[\Lambda \exp(i\Psi) - r\exp(i\psi)] \\ &\quad \left[\int_0^T \exp i(\theta_k + 2\phi t)dt + \int_0^T \exp(-i\theta_k)dt\right]. \end{aligned}$$

The first integral, a periodic function of T, is negligible in comparison with the second one. Then

$$\Delta w_k = (-\varepsilon T/2)\varrho_k[\Lambda \exp(i\Psi) - r\exp(i\psi)]\exp(-i\theta_k). \tag{A 3-1}$$

Since $\Delta \mathbf{y} = \sum(\Delta w_k)\mathbf{x}_k$ holds, substitution of (A3-1) in this summation gives

$$\begin{aligned} \Delta y &= \sum_{k=1}^{M} (-\varepsilon T/2)\varrho_k[\Lambda \exp(i\Psi) - r\exp(i\psi)]\exp(-i\theta_k) \\ &\quad 0.5[\varrho_k \exp(i\theta_k) + \varrho_k \exp(-i\theta_k)], \\ &= -(\varepsilon T/4)\left[\left(\sum_{k=1}^{M} \varrho_k^2\right)(\mathbf{y} - \mathbf{c}) \right. \\ &\quad \left. + \left(\sum_{k=1}^{M} \varrho_k^2 \exp(2i\theta_k)\right)(\mathbf{y} - \mathbf{c})^\dagger\right], \end{aligned}$$

which is equivalent to (3.28) with parameters μ and h in (3.29) and (3.30).

The substitution of (3.32) will transform (3.28) into

$$\Delta \mathbf{s} = -\mu(\mathbf{s} + \beta\mathbf{s}^\dagger), \tag{A3-2}$$

where $\Delta \mathbf{s}$ is a corresponding increment to $\Delta \mathbf{y}$. Let $\mathbf{y}[n]$ and $\mathbf{s}[n] = (\xi_n, \eta_n)$ be the values of $\mathbf{y}$ and $\mathbf{s}$ respectively after n times repetition of the learning process. Then it holds from (A3-2) that

$$\xi_n = [1 - \mu(1+\beta)]^n \xi_0$$

and

$$\eta_n = [1 - \mu(1-\beta)]^n \eta_n.$$

Elimination of exponents n with the help of approximation of $\log_e(1+\delta)$ by δ for small value of δ gives (3.34).

Acknowledgements. The author expresses his sincere thanks to Prof. Shun-ichi Amari, University of Tokyo, for his kind comments and encouragement in this research.

References

Albus, J.S.: A theory of cerebellar function. Math. Biosci. **10**, 25–61 (1971)

Amari, S.: Neural theory of association and concept formation. Biol. Cybern. **26**, 175–185 (1977)

Amari, S.: Topographic organization of nerve fields. Bull. Math. Biol. **42**, 339–364 (1980)

Amari, S., Takeuchi, A.: Mathematical theory on formation of category detecting nerve cells. Biol. Cybern. **29**, 127–136 (1978)

Andersson, G., Oscarsson, O.: Climbing fiber microzones in cerebellar vermis and their projection to different groups of cells in the lateral vestibular nucleus. Exp. Brain Res. **32**, 565–579 (1978)

Bell, C.C., Kawasaki, T.: Relations among climbing fiber responses nearby Purkinje cells. J. Neurophysiol. **35**, 155–169 (1972)

Calvert, T.W., Meno, F.: Neural systems modeling applied to the cerebellum. IEEE Trans. Syst. Man Cybern. SMC-**2**, 363–374 (1972)

Davies, P., Melvill Jones, G.: An adaptive neural model compatible with plastic changes induced in the human vestibulo-ocular reflex by prolonged optical reversal of vision. Brain Res. **103**, 546–550 (1976)

Denoth, F., Magherini, P.C., Pompeiano, O., Stanojevic, M.: Responses of Purkinje cells of cerebellar vermis to sinusoidal rotation of neck. J. Neurophysiol. **43**, 46–59 (1980)

Dufossé, M., Ito, M., Jastreboff, P.J., Miyashita, Y.: A neuronal correlate in rabbit's cerebellum to adaptive modification of the vestibulo-ocular reflex. Brain Res. **150**, 611–616 (1978)

Eccles, J.C.: An instruction-selective theory of learning in the cerebellar cortex. Brain Res. **127**, 327–352 (1977)

Eccles, J.C., Ito, M., Szentágothai, J.: The cerebellum as a neuronal machine. Berlin, Heidelberg, New York: Springer 1967

Fernandez, C., Goldberg, J.M.: Physiology of peripheral neurons innervating semicircular canals of the squirrel monkey. II. Response to sinusoidal stimulation and dynamics of peripheral vestibular system. J. Neurophysiol. **34**, 661–675 (1971)

Ghelarducci, B., Ito, M., Yagi, N.: Impulse discharges from flocculus Purkinje cells of alert rabbits during visual stimulation combined with horizontal head rotation. Brain Res. **87**, 66–72 (1975)

Gonshor, A., Melvill Jones, G.: Extreme vestibulo-ocular adaptation induced by prolonged optical reversal of vision. J. Physiol. (London) **256**, 381–414 (1976)

Groenewegen, H.J., Voogd, J.: The parasagittal zonation within the olivocerebellar projection. I. Climbing fiber destruction in the vermis of cat cerebellum. J. Comp. Neurol. **174**, 417–485 (1977)

Hassul, M., Daniels, P.D.: Cerebellar dynamics: The mossy fiber input. IEEE Trans. Biomed. Eng. BME-**24**, 449–456 (1977)

Higgins, D.C., Partridge, L.D., Glaser, G.H.: A transient cerebellar influence on stretch responses. J. Neurophysiol. **25**, 684–692 (1962)

Ito, M.: Neural design of the cerebellar motor control system. Brain Res. **40**, 81–84 (1972)

Ito, M.: Learning control mechanisms by the cerebellum investigated in the flocculo-vestibulo-ocular system. In: The Nervous System, Vol. 1, pp. 245–252. Tower, D.B.(ed). New York: Raven Press 1975

Ito, M.: Recent advances in cerebellar physiology and pathology. In: Advances in Neurology, Vol. 21, pp. 59–84. Kark, R.A., Rosenberg, R.N., Shut, L.J.(eds). New York: Raven Press 1978

Ito, M.: Is the cerebellum really a computer? Trends in Neurosciences **2**, 122–126 (1979)

Ito, M., Sakurai, M., Tongroach, P.: Climbing fibre induced depression of both mossy fibre responsiveness and glutamate sensitivity of cerebellar Purkinje cells. J. Physiol. **324**, 113–134 (1982)

Lange, W.: Regional differences in the distribution of Golgi cells in the cerebellar cortex of man and some other mammals. Cell Tiss. Res. **153**, 219–226 (1974)

Lisberger, S.G., Fuchs, A.F.: Role of primate flocculus during rapid behavioral modification of vestibuloocular reflex. II. Mossy fiber firing patterns during horizontal head rotation and eye movement. J. Neurophysiol. **41**, 764–777 (1978)

Marr, D.: A theory of cerebellar cortex. J. Physiol. (London) **202**, 437–470 (1969)

Miles, F.A., Fuller, J.H., Braitman, D.J., Dow, B.M.: Long-term adaptive changes in primate vestibuloocular reflex. III. Electrophysiological observations in flocculus of normal monkeys. J. Neurophysiol. **43**, 1437–1476 (1980)

Palkovits, M., Magyar, P., Szentagothai, J.: Quantitative histological analysis of the cerebellar cortex in the cat. I. Number and arrangement in space of the Purkinje cells. Brain Res. **32**, 1–13 (1971A)

Palkovits, M., Magyar, P., Szentagothai, J.: Quantitative histological analysis of the cerebellar cortex in the cat. II. Cell numbers and densities in the granular layer. Brain Res. **32**, 15–30 (1971B)

Sakrison, D.J.: Iterative design of optimum filters for non mean-square-error performance criteria. IEEE Trans. Inform. Theor. IT-**3**, 161–167 (1963)

Takahashi, Y., Rabins, M.J., Auslander, D.M.: Control and dynamic systems. Massachusetts: Addison-Wesley 1970

Tsukahara, N., Kiyohara, T., Ijichi, Y.: The mode of cerebellar control of pupillary light reflex. Brain Res. **60**, 244–248 (1973)

Tsypkin, Ya.Z.: Adaptation and learning in automatic systems. In: Mathematics in Science and Engineering, Vol. 73. Bellman, R.(ed). New York: Academic Press 1971

Wasan, M.T.: Stochastic approximation. Cambridge: Cambridge University Press 1969

Widrow, B., Glover, J.R., McCool, J.M., Kaunitz, J., Williams, C.S., Hearn, R.H., Zeidler, J.R., Dong, E., Goodlin, R.C.: Adaptive noise cancelling: Principles and applications. Proc. IEEE **63**, 1692–1716 (1975)

Widrow, B., Mantey, P.E., Griffiths, L.J., Goode, B.B.: Adaptive antenna systems. Proc. IEEE. **55**, 2143–2159 (1967)

Received: June 21, 1982

M. Fujita
Nagasaki Institute of Applied Science
536 Aba-machi
Nagasaki-shi
Nagasaki, 851-01
Japan