INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND GENERAL

J. Phys. A: Math. Gen. **35** (2002) 2379–2394 PII: S0305-4470(02)28638-2

Training the integrate-and-fire model with the informax principle: I

Jianfeng Feng[1,3], **Hilary Buxton**[1], **and Yingchun Deng**[2]

[1] COGS, Sussex University, Brighton BN1 9QH, UK

[2] Department of Mathematics, Hunan Normal University, Changsha 410081, People's Republic of China

Received 7 September 2001, in final form 31 December 2001
Published 1 March 2002
Online at stacks.iop.org/JPhysA/35/2379

Abstract

In terms of the informax principle, and the input–output relationship of the integrate-and-fire (IF) model, IF neuron learning rules are developed. For supervised learning and with uniform weight of synapses (the theoretically tractable case), we show that the derived learning rule is stable and the stable state is unique. For unsupervised learning, within physiologically reasonable parameter regions, both long-term potentiation (LTP) and long-term depression (LTD) could happen when the inhibitory input is weak, but LTD cannot be observed when inhibitory input is strong enough. When both LTP and LTD occur, LTD is observable when the output of the postsynaptic neuron is faster than pre-synaptic inputs, otherwise LTP is observable, as observed in recent experiments. Learning rules of general cases are also studied and numerical examples show that the derived learning rule tends to equalize the contribution of different inputs to the output firing rates.

PACS numbers: 87.18.Sn, 87.19.La, 05.10.Gg, 05.40.−a

1. Introduction

Learning or synaptic plasticity is of vital importance for biological systems [1]. In the present paper, we develop a learning rule, which is applicable to solving engineering problems [12] and is based upon (biophysical) models of a cell. The learning rule is derived under the principle of the maximization of the mutual information of input–output, which has been proposed and widely used in artificial neuron networks [2, 4, 18]. Due to recent developments in modelling single neurons, we know exactly the input–output relationship of some neuron models such as the integrate-and-fire (IF) model [27] and IF-FHN model [10] etc. Combining these two approaches together, we are able to develop learning rules relying on the input–output relationship of a neuron.

We first consider an ideal case where all synaptic strengths are identical. For supervised learning, by which we mean that the input and output firing rates of a neuron are fixed,

[3] http://www.cogs.susx.ac.uk/users/jianfeng

we prove that there are stable states for the derived learning rule. We also show that the stable state is unique. We then go further to consider unsupervised learning where the output firing rate is a function of inputs. We show that both long-term potentiation (LTP) and long-term depression (LTD) could occur, but LTD tends to be observable only when the inhibitory input is weak. This is reasonable since with LTD the total input to the neuron is reduced and therefore it is in a way equivalent to increasing inhibitory inputs. For the situation where LTP and LTD are both observable, there is a critical point of the efferent firing rate. Above it, which implies that the efferent firing rate is faster than input firing rates, LTD happens. Otherwise LTP occurs. This is in general agreement with recent experimental data [5, 19, 28] which show that when the postsynaptic neuron fires faster than the pre-synaptic neuron, LTD is observed, otherwise LTP happens. Very different from the previous applications of the informax principle [4], where the anti-Hebbian learning rules are found, the derived learning rules are coincident with recent experimental data.

By a simple combination of the informax principle and the input–output relationship of the IF model, some interesting and novel conclusions are obtained. Although we confine ourselves to the IF model, the method developed here can be applied to any neuron provided that we know its input–output relationship. In conclusion, we expect that the method presented here is fundamental and could open up many new and interesting problems for further studies.

2. Informax principle

We very briefly review some results on maximizing mutual information between the input and output of a system and refer the reader to [4] for details. For a given input X and output Y, the mutual information $M(Y, X)$ is defined by

$$M(Y, X) = H(Y) - H(Y|X)$$

where $H(Y)$ is the entropy of the output Y and $H(Y|X)$ is the conditional entropy. Under the assumption that the mapping between X and Y is deterministic, then maximizing the mutual information is reduced to maximizing the entropy $H(Y)$ (see section 8), as pointed out in [4]. Suppose that the output (firing frequency or interspike intervals) y (realization of Y) of a neuron is a function of input rate x (realization of X), with synaptic weights w, then the learning rule under the informax principle is to maximize $-\langle \log f(y)\rangle$ where $f(y)$ is the distribution density of y. Equivalently we have (equation (6) in [4])

$$\dot{w} \propto \left(\frac{\partial(-\langle \log f(y)\rangle)}{\partial w}\right) = \left(\frac{\partial y}{\partial x}\right)^{-1} \frac{\partial}{\partial w}\left(\frac{\partial y}{\partial x}\right) \tag{1}$$

which is the starting point of our development below.

3. The IF model

Suppose that a cell receives excitatory postsynaptic potentials (EPSPs) at p synapses and inhibitory postsynaptic potentials (IPSPs) at q inhibitory synapses. When the membrane potential V_t is between the resting potential V_{rest} and the threshold V_{thresh}, it is given by

$$\mathrm{d}V_t = -L(V_t - V_{rest})\,\mathrm{d}t + \mathrm{d}\bar{I}_{syn}(t) \tag{2}$$

where L is the decay rate and synaptic inputs

$$\bar{I}_{syn}(t) = \sum_{i=1}^{p} w_i^E E_i(t) - \sum_{j=1}^{q} w_j^I I_j(t) \tag{3}$$

with $E_i(t)$, $I_i(t)$ as Poisson processes with rate λ_i^E and λ_i^I respectively, $w_i^E > 0$, $w_j^I > 0$ being the magnitude of each EPSP and IPSP. Once V_t crosses V_{thresh} from below, a spike is generated and V_t is reset to V_{rest}. This model is termed the IF model [7, 8, 27].

Here we use the usual approximation to approximate the IF models, or more exactly the synaptic inputs of the models, i.e. the jump process in equation (3) is replaced by a diffusion process [27]. We do not check the approximation accuracy since it has been done by many authors [21, 27].

The input now reads

$$E_i(t) \sim \lambda_i^E t + \sqrt{\lambda_i^E} B_i^E(t)$$

and similarly

$$I_i(t) \sim \lambda_i^I t + \sqrt{\lambda_i^I} B_i^I(t)$$

where $B_i^E(t)$ and $B_i^I(t)$ are standard Brownian motions. Therefore the IF model can be approximated by

$$\mathrm{d}v_t = -L(v_t - V_{rest})\,\mathrm{d}t + \mathrm{d}\bar{i}_{syn}(t)$$

where

$$\bar{i}_{syn}(t) = \sum_{i=1}^{p} w_i^E \lambda_i^E t - \sum_{j=1}^{q} w_j^I \lambda_j^I t + \sum_{i=1}^{p} \sqrt{(w_i^E)^2 \lambda_i^E} B_i^E(t) - \sum_{j=1}^{q} \sqrt{(w_i^I)^2 \lambda_i^I} B_j^I(t). \tag{4}$$

Since the summation of Brownian motions is again a Brownian motion we can rewrite the equation above as follows:

$$\bar{i}_{syn}(t) = \mu t + \sigma B(t) \tag{5}$$

where $B(t)$ is a standard Brownian motion

$$\mu = \sum_{i=1}^{p} \lambda_i^E w_i^E - \sum_{j=1}^{q} \lambda_j^I w_j^I \qquad \sigma^2 = \sum_{i=1}^{p} \lambda_i^E (w_i^E)^2 + \sum_{j=1}^{q} \lambda_j^I (w_j^I)^2. \tag{6}$$

In the following we assume that $p = q$, $\lambda_j^I = r\lambda_j^E$ for $r \in [0, 1]$. Therefore when $r = 0$ the cell receives purely excitatory input and when $r = 1$ its inputs are exactly balanced. The interspike interval (ISI) of efferent spikes is

$$T(r) = \inf\{t : V_t \geqslant V_{thresh}\}.$$

We only consider the case of rate coding for the reason that a rigorous input–output relationship of firing rates is known for the IF model. By rate coding, we mean that the information is carried by the firing rates of a neuron. It is well known in the literature that the input–output relationship of a neuron takes a sigmoidal form and this is the basis of neural computations developed in the past decades. The input–output relationship of an IF model (see figure 1) takes a sigmoidal function as well (not surprising at all), but it depends not only on the mean of inputs, but also on the variance of inputs. The latter feature enables us to derive novel learning rules which, to the best of our knowledge, have not been reported in the literature and exhibit some intriguing phenomena. The importance that a neuron might use higher-order statistics to compute was recognized early in the literature (see for example [3]).

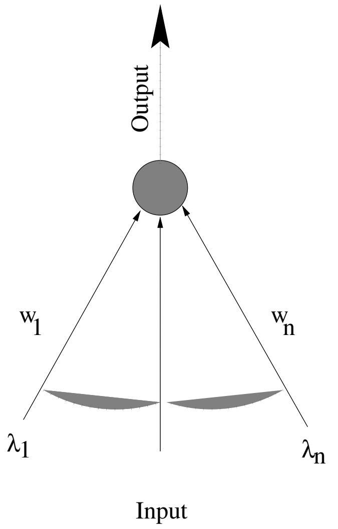

Figure 1. A schematic plot of the input and output relationship of the IF model. $w_1, w_2, \ldots, w_n$ are synaptic weights, $\lambda_1, \lambda_2, \ldots, \lambda_n$ are input rates.

4. Uniform synapse case

We first assume that $w_i^E = w_i^I = w, i = 1, \ldots, p, \sum_i \lambda_i^E = \lambda$. The assumption of $w_i = w$ is certainly not true and only of theoretical interest. Nevertheless, under the circumstances, we are able to carry out a rigorous study on the learning dynamics and gain insights into the general case, discussed in the following sections (see figure 1).

Let $\langle T(r)\rangle$ be the mean of the interspike intervals $T(r)$. For the IF model we have [10]

$$\langle T(r)\rangle = \frac{2}{L}\int_{A(V_{rest})}^{A(V_{thresh})} g(x)\,\mathrm{d}x \tag{7}$$

where

$$g(x) = \left[\exp(x^2)\int_{-\infty}^{x}\exp(-u^2)\,\mathrm{d}u\right]$$

and

$$A(V) = \frac{VL - \lambda(1-r)w}{w\sqrt{L\lambda(1+r)}}.$$

Therefore the output firing rate (Hz) is

$$\gamma = s(\lambda) = \frac{1000}{T_{ref} + \langle T(r)\rangle} \tag{8}$$

where T_{ref} is the refractory period. The input–output relationship given by equation (8) takes the form of a sigmoidal function, but is essentially different from the conventional expression of the input–output relationship. It depends on input not only through $w\lambda$ which appears in the conventional input–output relationship of a neuron, but also on $w\sqrt{\lambda}$ (the variance term). In fact in neuroscience it is traditional to exclusively consider $(1/\langle T\rangle)$–μ (output firing rate–input mean) curves. Recently [11] we have emphasized that to gain a complete picture of neuronal response properties, we should look at response surfaces, i.e. $(1/\langle T\rangle) - (\mu, \sigma)$.

Now we apply the informax principle to the input–output relationship of the IF model. To this end, we have

$$\frac{\partial\langle T(r)\rangle}{\partial\lambda} = -\frac{2}{L}g\left(\frac{V_{thresh}L - \lambda(1-r)w}{w\sqrt{\lambda L(1+r)}}\right)\frac{w(1-r)\lambda + V_{thresh}L}{2w\lambda\sqrt{\lambda(1+r)L}}$$

$$+\frac{2}{L}g\left(\frac{V_{rest}L-\lambda(1-r)w}{w\sqrt{\lambda L(1+r)}}\right)\frac{w(1-r)\lambda+V_{rest}L}{2w\lambda\sqrt{\lambda(1+r)L}}.$$

From now on we further assume that $V_{rest} = 0$ mV the equation above thus becomes

$$\frac{\partial\langle T(r)\rangle}{\partial\lambda} = -\frac{2}{L}g\left(\frac{V_{thresh}L-\lambda(1-r)w}{w\sqrt{\lambda L(1+r)}}\right)\frac{w(1-r)\lambda+V_{thresh}L}{2w\lambda\sqrt{\lambda(1+r)L}}$$
$$+\frac{2}{L}g\left(\frac{-\lambda(1-r)}{\sqrt{\lambda L(1+r)}}\right)\frac{(1-r)}{2\sqrt{\lambda(1+r)L}}.$$

Note that the second term in the expression above is independent of w, learning independent.

Similarly for $\partial\langle T(r)\rangle/\partial w$ we have

$$\frac{\partial\langle T(r)\rangle}{\partial w} = -\frac{2}{L}g\left(\frac{V_{thresh}L-\lambda(1-r)w}{w\sqrt{\lambda L(1+r)}}\right)\frac{\sqrt{L}V_{thresh}}{w^2\sqrt{\lambda(1+r)}}.$$

For simplicity of calculation we introduce two new variables

$$U = \frac{V_{thresh}L-\lambda(1-r)w}{w\sqrt{\lambda L(1+r)}} \qquad V = \frac{V_{thresh}L+\lambda(1-r)w}{w\sqrt{\lambda L(1+r)}}$$

and note the relation that $g'(x) = 2xg(x) + 1$; we obtain

$$\frac{\partial}{\partial w}\left(\frac{\partial\langle T(r)\rangle}{\partial\lambda}\right) = \frac{2}{L}[(2Ug(U)+1)V+g(U)]\frac{V_{thresh}\sqrt{L}}{2w^2\lambda\sqrt{\lambda(1+r)}}.$$

According to equation (1) define

$$l(w) = \left(\frac{\partial\gamma}{\partial\lambda}\right)^{-1}\frac{\partial}{\partial w}\left(\frac{\partial\gamma}{\partial\lambda}\right); \tag{9}$$

we have

$$l(w) = -\frac{2(\partial\langle T(r)\rangle/\partial w)\partial\langle T(r)\rangle/\partial\lambda + (\partial/\partial w)\big(\partial\langle T(r)\rangle/\partial\lambda\big)\cdot(T_{ref}+\langle T(r)\rangle)}{(T_{ref}+\langle T(r)\rangle)\cdot\partial\langle T(r)\rangle/\partial\lambda}$$
$$= -\frac{\gamma\partial\langle T(r)\rangle/\partial w}{500} + \frac{(\partial/\partial w)\big(\partial\langle T(r)\rangle/\partial\lambda\big)}{\partial\langle T(r)\rangle/\partial\lambda} \tag{10}$$
$$= \frac{[(2Ug(U)+1)V+g(U)]V_{thresh}\sqrt{L}/w^2\lambda\sqrt{\lambda(1+r)}}{-g(U)V+g\big(-\lambda(1-r)/\sqrt{\lambda L(1+r)}\big)(1-r)/\sqrt{\lambda(1+r)L}}$$
$$+\frac{V_{thresh}}{250\sqrt{(1+r)L}}\frac{\gamma g(U)}{w^2\sqrt{\lambda}}. \tag{11}$$

Looking at the second term in the learning rule above (equation (11)), we see that it gives us an anti-Hebb learning rule, as found in terms of the conventional input–output neuron relationship. Note that the anti-Hebb learning rule developed here takes the form of $output/input$ rather than $-(output)\cdot(input)$. We shall see the implication of the form later on.

We are particularly interested in two cases: $r = 1$ (exactly balanced inputs) and $r = 0$ (purely excitatory inputs). These two cases have been intensively studied in the literature in recent years. A biological neuron system is composed of excitatory and inhibitory neurons. A recent hypothesis [24] on single neuron modelling is that a neuron receives exactly balanced excitatory and inhibitory inputs, namely $r = 1$. With exactly balanced inputs, it is observed that the neuronal output is in general more irregular than that with purely excitatory inputs.

- Exactly balanced case: $r = 1$. The learning rule of equation (11) turns out to be

$$\begin{aligned} l(w) &= \frac{V_{thresh}}{250\sqrt{2}}\frac{\gamma g(U)}{w^2\sqrt{L\lambda}} - \left[2U + \frac{1}{g(U)} + \frac{1}{V}\right]\frac{V_{thresh}\sqrt{L}}{w^2\lambda\sqrt{2\lambda}} \\ &= -\left[2\frac{V_{thresh}L}{w\sqrt{2\lambda L}} + \left(g\left(\frac{V_{thresh}L}{w\sqrt{2\lambda L}}\right)\right)^{-1} + \frac{w\sqrt{2\lambda L}}{V_{thresh}L}\right]\frac{V_{thresh}\sqrt{L}}{w^2\lambda\sqrt{2\lambda}} \\ &\quad + \frac{V_{thresh}}{250\sqrt{2}w^2\sqrt{L}}g\left(\frac{V_{thresh}L}{w\sqrt{2\lambda L}}\right)\frac{\gamma}{\sqrt{\lambda}}. \end{aligned} \tag{12}$$

The first term is negative and the second is positive, depending both on inputs and outputs.

- Purely excitatory input case: $r = 0$.

$$l(w) = \frac{[(2Ug(U)+1)V + g(U)]V_{thresh}\sqrt{L}/w^2\lambda\sqrt{\lambda}}{-g(U)V + g\left(-\lambda/\sqrt{\lambda L}\right)/\sqrt{\lambda L}} + \frac{V_{thresh}}{250}\frac{\gamma g(U)}{w^2\sqrt{L\lambda}}. \tag{13}$$

In order to understand the behaviour of $l(w)$, let us have a look of the case of $r = 1$. When w is small, the dominant term in equation (12) is

$$g\left(\frac{V_{thresh}L}{w\sqrt{2\lambda L}}\right) \sim \exp\left[\frac{V_{thresh}L}{w\sqrt{2\lambda L}}\right]^2$$

which rapidly diverges to positive infinity. Hence when the weight is small, it increases rapidly. When w is large enough,

$$g\left(\frac{V_{thresh}L}{w\sqrt{2\lambda L}}\right) \sim \exp\left[\frac{V_{thresh}L}{w\sqrt{2\lambda L}}\right]^2$$

tends to a finite constant and the leading term is

$$-\frac{w\sqrt{2\lambda L}}{V_{thresh}L}\frac{V_{thresh}\sqrt{L}}{w^2\lambda\sqrt{2\lambda}}$$

which implies that $l(w)$ is negative and tends to zero when w is large enough (see figure 2). The conclusion above is true for the general case of $r \in [0, 1]$. A simple implication of the observation above is that the derived learning rule is always stable and the stable state is positive. The learning rule of equation (11) automatically prevents the weight from changing its sign.

Theorem 1. *The derived learning rule is stable and the weight does not change its sign. Furthermore, the stable state is unique and is the solution of the following equation:*

$$\left[2U + \frac{1}{g(U)}\right]V + 1 = \frac{\gamma\lambda}{250L}\left[g(U)V - g\left(\frac{-\lambda(1-r)}{\sqrt{\lambda L(1+r)}}\right)\frac{(1-r)}{\sqrt{\lambda(1+r)L}}\right]. \tag{14}$$

We postpone the proof of the uniqueness to the appendix; equation (14) is directly from equation (11).

A comparison between the exactly balanced and purely excitatory input case is illuminating, as shown in figure 2. For fixed input rates, the stable weight is an increasing function of r. This is simply coincident with our intuition: the larger r, the weaker the mean input and so the larger the weight required. For the same reason the stable weight is also a decreasing function of L. Furthermore the learning rule is not symmetric: the weight tends to stay at a large value rather than a small value. The assumption that a neuron receives an exactly balanced input is interesting and is currently a hot topic both in theory and in experiments [24, 25]. From figure 2, $r = 1$ we could also see that in this case the stable

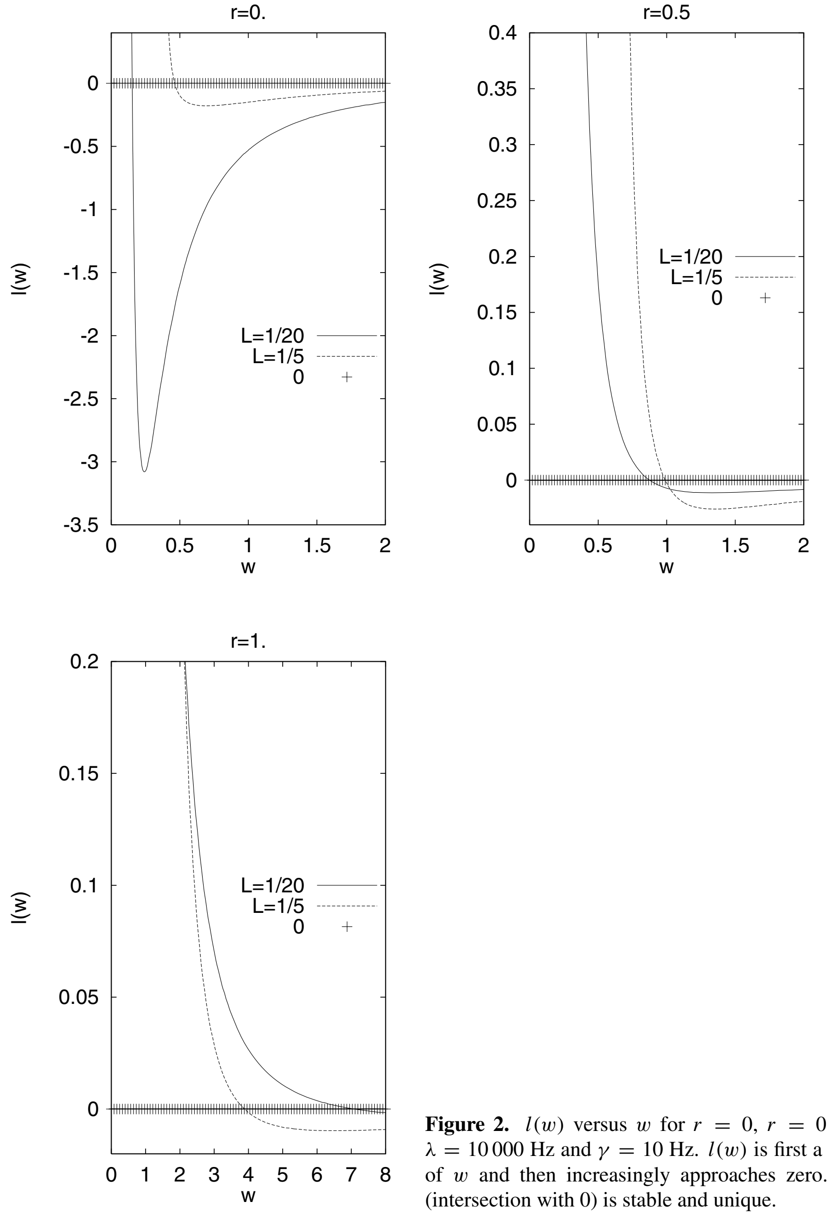

Figure 2. $l(w)$ versus w for $r = 0$, $r = 0.5$ and $r = 1$ with $\lambda = 10\,000$ Hz and $\gamma = 10$ Hz. $l(w)$ is first a decreasing function of w and then increasingly approaches zero. The stable point (intersection with 0) is stable and unique.

state of the weight is quite high. It is reported in the literature that the magnitude of EPSP is around 0.5 mV with a maximum of 2 mV [24].

The purpose of a learning rule is to find a set of weights w which could produce the given input and output firing rates. Does the learning rule presented here achieve the goal? From numerical results in the literature [9], we conclude that for the parameters given in figure 2 and $w > 0.5$, γ will be around 10 Hz, after adding an appropriate refractory period.

5. LTP versus LTD

In the previous section we considered the case of learning with fixed input and output firing rates. This is supervised learning. In this section we consider unsupervised learning. There is a large body of literature devoted to the topic, both theoretically and experimentally. Basically it is found that with a slow firing input, the synapse will gradually die away, but with a strong firing input, the synaptic weight will gradually increase.

For our purpose we rewrite the learning rule equation (11) as follows:

$$l(w) = \lambda\Phi(\gamma, \gamma_m)$$

where γ_m is the point with the property that

$$\Phi(\gamma_m, \gamma_m) = 0.$$

Now we fix input rates λ and consider the behaviour of $\Phi(\gamma, \gamma_m)$. Figure 3 clearly shows that when $r = 0$ and $\gamma < 59$ Hz, LTP occurs, i.e. the weight increases its value. However when $\gamma > 59$ Hz LTD happens, i.e. the weight decreases its value. The phenomenon we observe here is of extreme interest. From a purely theoretical approach we demonstrate that both LTP and LTD exist. Remembering that the summation of Poisson processes is again a Poisson process, we therefore could let $\lambda_i = \gamma_m$, i.e. $p = \lambda/\gamma_m$. We then have the following scenario. When the output firing rate is higher than the input firing rate, LTD happens, and when the output firing rate is lower than the input firing rate, LTP occurs. This learning also reminds us of the well known Bienenstock, Cooper and Munro (BCM) learning rule [15]. Quantitatively the BCM learning rule and equation (11) have similar behaviour: when the output firing rate is fast the LTP is introduced and otherwise the LTD happens.

If we suppose that λ and γ are instantaneous firing rates, then the results presented above are also coincident with recent experiments. Basically, figure 3 ($r = 0$) tells us that when the postsynaptic neuron fires a spike later than its pre-synaptic neuron, the synaptic strength increases; otherwise it decreases. In experiments on neocortical slices [19], hippocampal cells [5] in culture and *in vivo* studies of tadpole tectum [28], LTP occurs if pre-synaptic action potentials preceded post-synaptic firing by no more than about 50 ms. If pre-synaptic spikes follow post-synaptic action potentials, LTD rather than LTD is observed. Figure 4 shows $l(w)$ versus 1000/(input ISI) − 1000/(output ISI) with a fixed input ISI = 59 Hz (see for example, figure 1 in [6] for a comparison).

In recent years, there have been many research activities devoted to the topic of learning in time domains [13, 14]. In most of the learning rules used for theoretical study, the updating of dynamics is according to an *ad hoc* relationship [16,25]. Our results presented here provide us with a quantitative rule which shows both LTP and LTD.

Finally we want to emphasize that from figure 3 we know that LTD is observable only when a neuron receives relatively weak inhibitory inputs. If there are strong inhibitory inputs present, then only LTP is observed. This seems a reasonable phenomenon and we thus predicate that in a real neuronal system a similar situation could be true.

For the situations we considered in the previous two sections, an essential shortcoming of derived learning rules is the requirement of synchronously updating weights. In the next section, we propose an asynchronous learning rule (see [2, 17]).

6. General case

To overcome the shortcoming mentioned in the previous section, here we consider a local learning rule under the assumption that the ith synapse is capable of detecting its local field of

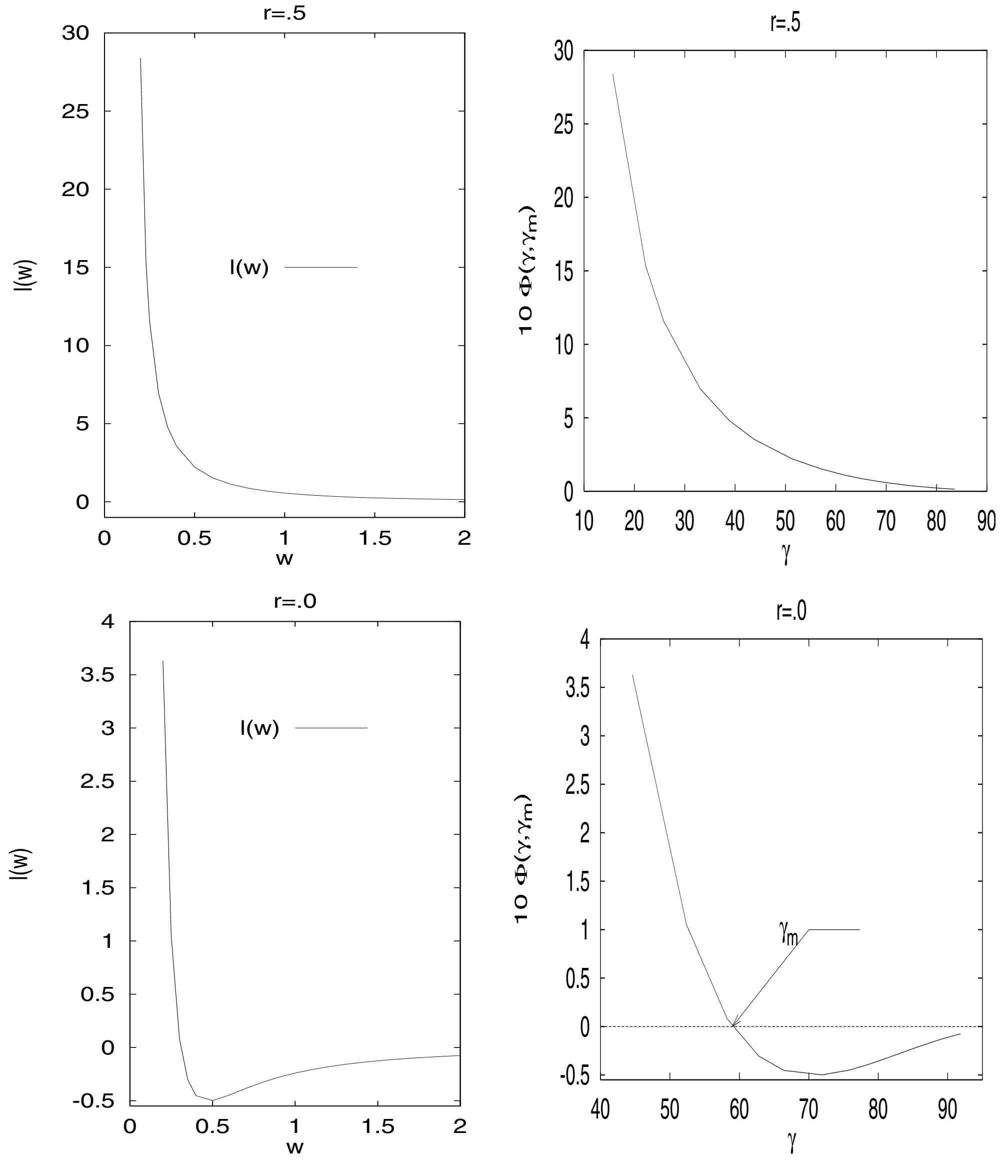

Figure 3. Output frequency is obtained by adding a refractory of 10 ms, $L = 1/20$, $\lambda = 10\,000$ Hz.

mean $\mu_i = \sum_{j\neq i} w_j\lambda_j$ and variance $\sigma_i^2 = \sum_{j\neq i} w_j^2\lambda_j$ of inputs. Under this assumption, we can then maximize the local entropy, as developed below. In fact this formulation also allows the systems to update asynchronously, rather than synchronously, as in most learning rules derived under the informax principle.

Remember that the mean firing time is given by

$$\langle T(r)\rangle = \frac{2}{L}\int_{B(V_{rest})}^{B(V_{thresh})} g(x)\,\mathrm{d}x \tag{15}$$

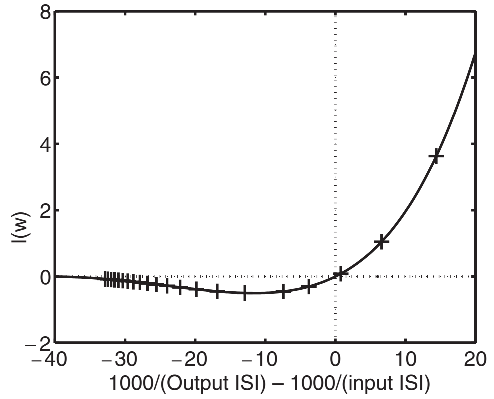

Figure 4. $l(w)$ versus 1000/(input ISI) − 1000/(output ISI) with $r = 0$ (see figure 3).

where

$$B(V) = \frac{VL - (w_1\lambda_1 + \mu_1)(1-r)}{\sqrt{(w_1^2\lambda_1 + \sigma_1^2)L(1+r)}}.$$

Defining

$$\Sigma_1^2 = (w_1^2\lambda_1 + \sigma_1^2)L(1+r)$$
$$\Theta_1 = (w_1\lambda_1 + \mu_1)(1-r)$$

and

$$\xi(x) = \frac{2w_1(1-r)\Sigma_1^2 + w_1^2 L(1+r)(xL - \Theta_1)}{2\Sigma_1^3}$$
$$\eta(x) = \frac{\lambda_1(1-r)\Sigma_1^2 + (xL - \Theta_1)w_1\lambda_1 L(1+r)}{\Sigma_1^3}$$
$$\zeta(x) = \frac{2(1-r)\Sigma_1^2 + 3w_1^2\lambda_1 L(1-r^2) + 2w_1 L(1+r)(xL - \Theta_1)}{2\Sigma_1^3} - \frac{3w_1\lambda_1 L(1+r)[2w_1(1-r)\Sigma_1^2 + w_1^2 L(1+r)(xL - \Theta_1)]}{2\Sigma_1^5}$$

we arrive at

$$\begin{aligned}\frac{\partial\langle T(r)\rangle}{\partial\lambda_1} &= -\frac{2}{L}g\left(\frac{V_{thresh}L - \Theta_1}{\Sigma_1}\right)\xi(V_{thresh}) + \frac{2}{L}g\left(-\frac{\Theta_1}{\Sigma_1}\right)\xi(0)\\ \frac{\partial\langle T(r)\rangle}{\partial w_1} &= -\frac{2}{L}g\left(\frac{V_{thresh}L - \Theta_1}{\Sigma_1}\right)\eta(V_{thresh}) + \frac{2}{L}g\left(-\frac{\Theta_1}{\Sigma_1}\right)\eta(0)\end{aligned} \tag{16}$$

and

$$\begin{aligned}\frac{\partial}{\partial w_1}\left(\frac{\partial\langle T(r)\rangle}{\partial\lambda_1}\right) = &-\frac{2}{L}g\left(\frac{V_{thresh}L - \Theta_1}{\Sigma_1}\right)\zeta(V_{thresh})\\ &+\frac{2}{L}\left[2g\left(\frac{V_{thresh}L - \Theta_1}{\Sigma_1}\right)\frac{V_{thresh}L - \Theta_1}{\Sigma_1} + 1\right]\xi(V_{thresh})\eta(V_{thresh})\\ &+\frac{2}{L}g\left(\frac{-\Theta_1}{\Sigma_1}\right)\zeta(0) - \frac{2}{L}\left[2g\left(\frac{-\Theta_1}{\Sigma_1}\right)\frac{-\Theta_1}{\Sigma_1} + 1\right]\xi(0)\eta(0).\end{aligned} \tag{17}$$

When $\Theta_1 = w_1\lambda_1(1-r)$ and $\Sigma_1^2 = w_1^2\lambda_1(1+r)$ equations (16) and (17) are simply reduced to the corresponding equations in the previous subsection with $w = w_1, \lambda = \lambda_1$. Combining equations (16), (17) with (10), we obtain a novel learning rule based upon the IF model.

The learning rule presented here is too complex to be explored theoretically; nevertheless, we could simulate it numerically, as presented in the next section.

7. Numerical results

To see the effect of the learning rule presented in the previous section, we simulate the learning rule with the following parameters: $p = 500$, $\gamma = 20$ Hz, $\lambda_i = 30$ Hz for $i = 1, \ldots, 250$, $\lambda_i = 10$ Hz for $i = 250, \ldots, 500$, $r \in [0, 1]$ and $L = 1/20$. The initial weights are random variables uniformly distributed in [0, 1] and the step size is 0.0001.

Figure 5 shows the results after ten and 2000 steps of learning. After ten steps of learning we see that weights are almost uniformly distributed within [0, 1]. Nevertheless, after 2000 steps of learning, the picture is then changed. For $r = 0$ all weights are now greater than 0.5. The mean and standard variation of the first 250 weights are 0.826 409 and 0.166 225 respectively; those for the second 250 weights are 0.806 036 and 0.152 571. Hence there is a small difference between the first 250 weights (receiving stronger inputs) and the second 250 weights. The mean and standard variation of the total 500 synapses are 0.816 222 and 0.159 869. Having another ten steps of learning still increases the weight, but very slightly (with total mean and standard variation of 0.817 37 and 0.159 63). A similar phenomenon is observed for $r = 1$, although now the weights converge to a higher value but a small standard deviation (with a mean and standard deviation of 1.070 18 and 0.128 412).

Now we simulate the learning rule of unsupervised learning (figure 6). The most interesting thing to note is that the weights are no longer uniform when $r = 0$. For neurons with weaker inputs, their weights are greater than those with stronger inputs. In other words, the derived learning rule here assigns strong weights to weak inputs and therefore equalizes their contribution to the output firing rates. Another general feature is that after learning, the range of weights is much narrower than before learning. Again it equalizes the contribution of different synaptic inputs to the output.

8. Discussion

We have presented a theoretical approach to derive novel learning rules based upon spiking neurons. In particular, for the IF model and both the supervised learning and unsupervised learning, we have proved that when the synapses are uniform, the learning rule obtained under the informax principle is stable. We have also demonstrated that the stable state is unique. For unsupervised learning, we conclude that when the inhibitory input is weak both LTP and LTD are observable. Most interestingly, the derived learning rule quantitatively agrees with recently experimental data. For the general case, supervised learning and unsupervised learning have been investigated and numerical results are included.

The approach presented here is quite general. In principle, once we know the input–output relationship of a neuron, we could obtain the learning rule of the neuron. As we have mentioned in the introduction, we have accumulated many results on the input–output relationship of a single neuron, for example on the IF-FHN model [10].

All results presented here are closer to a computational neuroscience approach, rather than solving practical problems. We shall present the application of learning rules to practical tasks elsewhere [12].

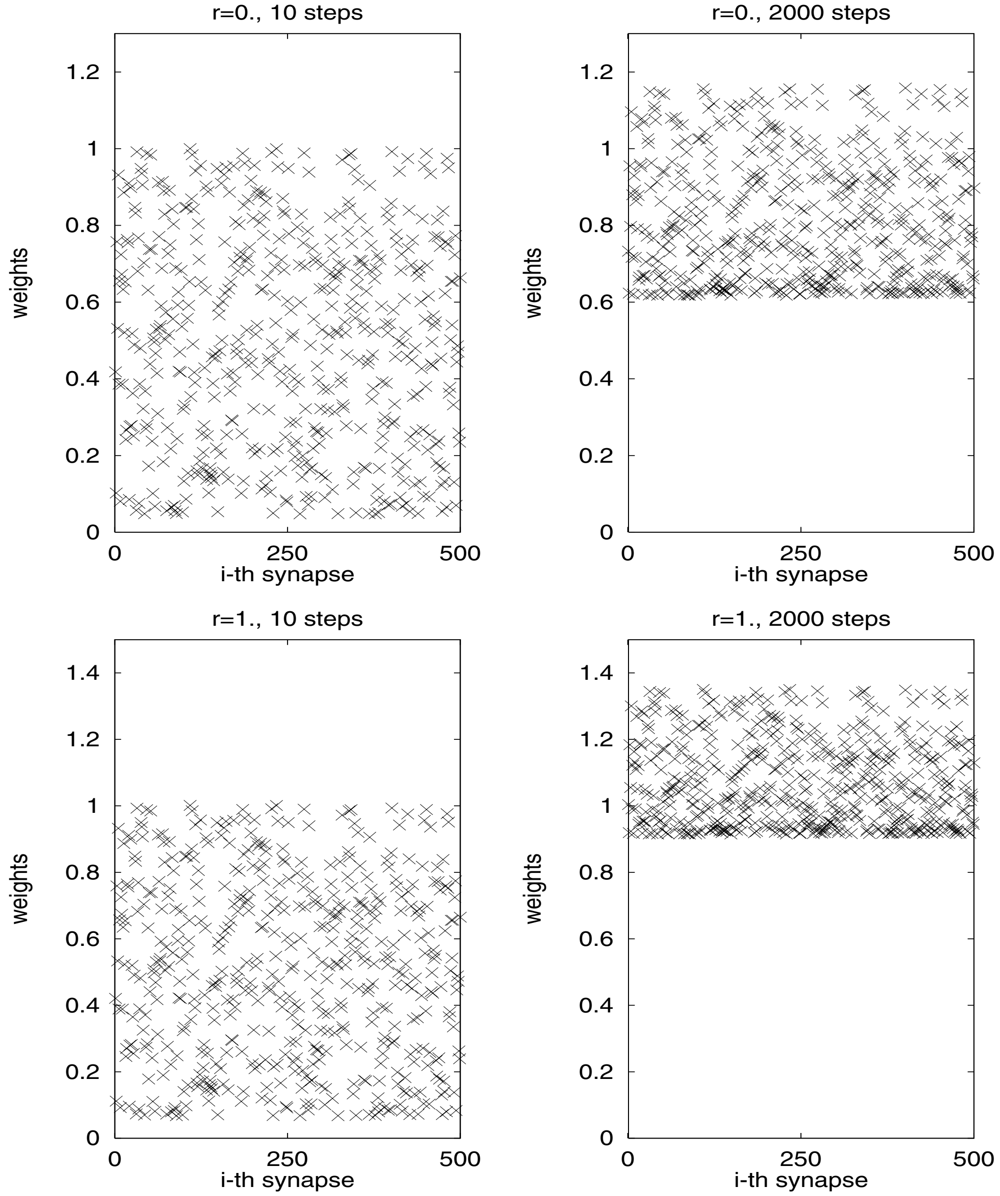

Figure 5. Synaptic weights after various steps of iteration for $r = 0$ and 1.

One of the drawbacks of the current work is that we focus only on the rate coding assumption, although it remains a widely accepted hypothesis in neuroscience. It would be very interesting to carry out a similar study for the time coding assumption [13]. The essential difficulty in carrying out such a study lies in the fact that it is in general not easy to obtain an analytical input–output relationship of the ISI distribution. In fact, despite many years of research endeavours, we know that such a formula is lacking [10]. For the IF-FHN model [10], we have demonstrated that Kramer's formula [22] gives a reasonable approximation to the distribution of interspike intervals. Hence it is possible to apply the ideas in the current paper to the IF-FHN model, under the time coding assumption. It would also be intriguing to

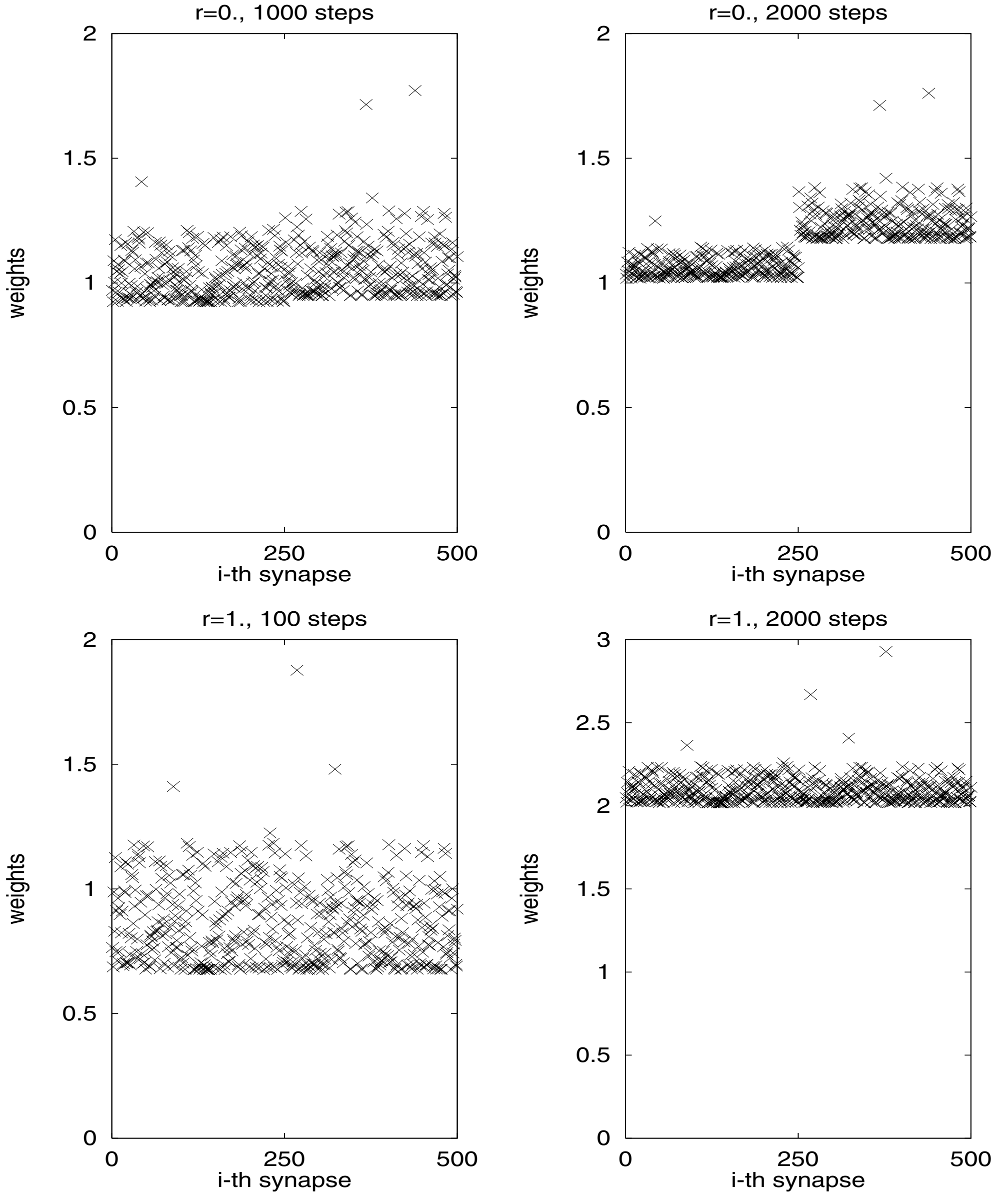

Figure 6. Synaptic weights after various steps of iteration for $r = 0$ and 1.

consider the case of correlated inputs, which is currently intensively investigated [10, 23, 26]. Another closely related issue is that the entropy maximization rather than a genuine information maximization is applied, as pointed out in section 2. If we simply assume that $Y = G(X) + N$ (see section 2), where G is an invertible transformation and N is additive noise on the outputs, for the IF model we have $G = \langle T(r) \rangle$. In this case we see that [20] $H(Y|X) = H(N)$ independent of w. Hence maximizing the mutual information is equivalent to maximizing the entropy. Nevertheless, in general changing weights and the ratio between inhibitory inputs and excitatory inputs will both lead to a change of the output distribution of the interspike

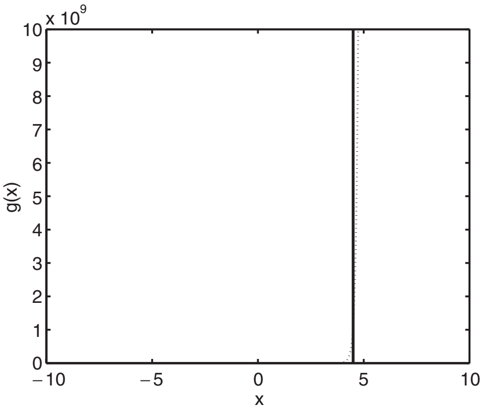

Figure A.1. $g(x)I_{\{g(x)<10^{10}\}} + 10^{10}I_{\{g(x)\geqslant 10^{10}\}}$ versus x (dotted line), and $10^{10}I_{\{x\geqslant 4.3\}}$ versus x (thick vertical line).

intervals (see for example, [10]) and therefore $H(Y|X)$ depends on w. To develop an exact learning rule based upon the information maximization then requires an exact expression of the distribution of the interspike intervals.

Acknowledgments

We are grateful to the referees for their constructive comments on an early version of the paper. JF was partially supported by a grant from EPSRC(GR/R54569) and a grant from the Royal Society.

Appendix

We prove the remaining conclusions in theorem 1. As before, we only consider the case of $r = 1$. The general case can be treated similarly. We first check that there is only one point of $w > 0$ at which $l'(w) = 0$. Since we are only interested in the roots of $l'(w) = 0$, we only need to consider

$$\begin{aligned}\bar{l}(w) &= -\left[2\frac{c}{w} + \left(g\left(\frac{c}{w}\right)\right)^{-1} + \frac{w}{c}\right]\frac{\sqrt{L}}{w^2\lambda} + \frac{\gamma}{250w^2\sqrt{L}}g\left(\frac{c}{w}\right) \\ &= -\left[2\frac{c}{w} + \left(g\left(\frac{c}{w}\right)\right)^{-1} + \frac{w}{c}\right]\frac{\sqrt{L}}{w^2\lambda} + \frac{4}{w^2\sqrt{L}(T_{ref} + \langle T(1)\rangle)}g\left(\frac{c}{w}\right) \qquad (18)\end{aligned}$$

where $V_{thresh}L/\sqrt{2\lambda L} = c$. The derivative of $\bar{l}(w)$ is then given by

$$\begin{aligned}&\frac{2\sqrt{L}g(c/w)/\lambda w^3 + g'(c/w)c\sqrt{L}/w^4\lambda}{g^2(c/w)} + \frac{6c\sqrt{L}}{w^4\lambda} + \frac{\sqrt{L}}{c\lambda w^2} \\ &\quad - \frac{4g'(c/w)c\sqrt{L}(T_{ref} + \langle T(1)\rangle) + 4g(c/w)[2w\sqrt{L}(T_{ref} + \langle T(1)\rangle) - c\sqrt{L}g(c/w)]}{w^4L(T_{ref} + \langle T(1)\rangle)^2}. \qquad (19)\end{aligned}$$

A key observation (see figure A.1) in our proof is that we can further approximate g by

$$g(x) \sim cI_{\{x\leqslant 4.3\}} + \infty I_{\{x\geqslant 4.3\}}$$

where $I_A(x) = 1$ if $x \in A$ and 0 otherwise, i.e. the indicator function. We first consider the case that

$$c/w < 4.3.$$

In this case we see that the first term in equation (19) is a higher-order term and so we can simplify equation (19) as

$$-\frac{4g'(c/w)c\sqrt{L}(T_{ref}+\langle T(1)\rangle)+4g(c/w)[2w\sqrt{L}(T_{ref}+\langle T(1)\rangle)-c\sqrt{L}g(c/w)]}{w^4L(T_{ref}+\langle T(1)\rangle)^2} \\ +\frac{6c\sqrt{L}}{w^4\lambda}+\frac{\sqrt{L}}{c\lambda w^2}. \tag{20}$$

Note that

$$\frac{g(c/w)}{T_{ref}+\langle T(1)\rangle}\sim\frac{g'(c/w)}{g(c/w)}\sim 2c/w;$$

we then have

$$(\bar{l}(w))'\sim\frac{6c\sqrt{L}}{w^4\lambda}+\frac{\sqrt{L}}{c\lambda w^2}-\frac{4c\sqrt{L}}{w^4L}\left(\frac{2c}{w}\right)^2-\frac{8w\sqrt{L}}{w^4L}\frac{2c}{w}+\frac{4c\sqrt{L}}{w^4L}\left(\frac{2c}{w}\right)^2.$$

Therefore the only possible solution of $(\bar{l}(w))'=0$ is

$$w=\frac{L\lambda}{c\lambda(16c\lambda-4cL)}.$$

When $c/w > 4.3$, using the relationship that $g(x)\sim x$, we can conclude that there is no solution for $(\bar{l}(w))'=0$. The results above, together with the fact that $l(w)\to\infty$ as $w\to 0$ and $l(w)\to 0$ with $l(w)<0$ as $w\to\infty$, imply the conclusion.

References

[1] Albright T D, Jessell T M, Kandel E R and Posner M I 2000 Neural science: a century of progress and the mysteries that remain *Cell* **100** s1–55

[2] Amari S 1999 Natural gradient learning for over- and under-complete bases in ICA *Neural Comput.* **11** 1875–83

[3] Barlow H 1986 Perception: what quantitative laws govern the acquisition of knowledge from the senses? *Functions of the Brain* ed C Coen (Oxford: Clarendon) pp 11–43

[4] Bell A J and Sejnowski T J 1995 An information maximization approach to blind separation and blind deconvolution *Neural Comput.* **7** 1129–59

[5] Bi G-q and Poo M-m 1998 Activity-induced synaptic modifications in hippocampal culture: dependence on spike timing, synaptic strength and cell type *J. Neurosci.* **18** 10 464–72

[6] Bi G-Q and Poo M-M 2001 Synaptic modification by correlated activity: Hebb's postulate revisited *Annu. Rev. Neurosci.* **24** 139–66

[7] Brown D, Feng J and Feerick S 1999 Variability of firing of Hodgkin–Huxley and FitzHugh–Nagumo neurons with stochastic synaptic input *Phys. Rev. Lett.* **82** 4731–4

[8] Feng J 1997 Behaviours of spike output jitter in the integrate-and-fire model *Phys. Rev. Lett.* **79** 4505–8

[9] Feng J and Brown D 1998 Impact of temporal variation and the balance between excitation and inhibition on the output of the perfect integrate-and-fire model *Biol. Cybern.* **78** 369–76

[10] Feng J 2001 Is the integrate-and-fire model good enough?—a review *Neural Networks* **14** 955–75

[11] Feng J, Brown D and Li G 2000 Synchronization due to common pulsed input in Stein's model *Phys. Rev.* E **61** 2987–95

[12] Feng J and Buxton H 2002 Training integrate-and-fire models with the informax principle II *IEEET. on Neural Networks* submitted

[13] Gerstner W, Kreiter A K, Markram H and Herz A V M 1997 Neural codes: firing rates and beyond *Proc. Natl Acad. Sci. USA* **94** 12 740–1

[14] Hopfield J J and Herz A V M 1995 Rapid local synchronization of action-potentials-toward computation with coupled integrate-and-fire neurons *Proc. Natl Acad. Sci. USA* **92** 6655–62

[15] Koch C 1999 *Biophysics of Computation* (Oxford: Oxford University Press)

[16] Kempter R, Gerstner W and van Hemmen J L 1999 Hebbian learning and spiking neurons *Phys. Rev.* E **59** 4498–514

[17] Lewicki M S and Sejnowski T J 2000 Learning overcomplete representations *Neural Comput.* **12** 337–65

[18] Linsker R 1989 An application of the principle of maximum information preservation to linear systems *Advances in Neural Information Processing Systems 1* ed D S Touretzky (San Mateo, CA: Morgan Kauffman)

[19] Markram H, Lubke J, Frotscher M and Sakmann B 1997 Regulation of synaptic efficacy by coincidence of post-synaptic APs and EPSPs *Science* **275** 213–15

[20] Nadal and Parga 1994 Nonlinear neurons in the low noise limit: a factorial code maximises information transfer *Network* **5** 565–81

[21] Ricciardi L M and Sato S 1990 Diffusion process and first-passage-times problems *Lectures in Applied Mathematics and Informatics* ed L M Ricciardi (Manchester: Manchester University Press)

[22] Risken S 1989 *The Fokker–Planck Equation* (Berlin: Springer)

[23] Salinas E and Sejnowski T 2001 Correlated neuronal activity and the flow of neural information *Nature Rev. Neurosci.* **2** 539–50

[24] Shadlen M N and Newsome W T 1994 Noise, neural codes and cortical organization *Curr. Opin. Neurobiol.* **4** 569–79

[25] Song S, Miller K D and Abbott L F 2000 Competitive Hebbian learning through spike-timing-dependent synaptic plasticity *Nature Neurosci.* **3** 919–26

[26] Stevens C F and Zador A M 1998 Input synchrony and the irregular firing of cortical neurons *Nature Neurosci.* **1** 210–17

[27] Tuckwell H C 1988 *Introduction to Theoretical Neurobiology* vol 2 (Cambridge: Cambridge University Press)

[28] Zhange L I, Tao H W, Holt C E, Harris W A and Poo M-M 1998 A critical window for cooperation and competition among developing retinotectal synapses *Nature* **395** 37–44