18 Grossberg Network

Objectives

This chapter is a continuation of our discussion of associative and competitive learning algorithms in Chapters 15 and 16. The Grossberg network described in this chapter is a self-organizing continuous-time competitive network. This will be the first time we have considered continuous-time recurrent networks, and we will introduce concepts here that will be further explored in Chapters 20 and 21. This Grossberg network is also the foundation for the adaptive resonance theory (ART) networks that we will present in Chapter 19.

We will begin with a discussion of the biological motivation for the Grossberg network: the human visual system. Although we will not cover this material in any depth, the Grossberg networks are so heavily influenced by biology that it is difficult to discuss his networks without putting them in their biological context. It is also important to note that biology provided the original inspirations for the field of artificial neural networks, and we should continue to look for inspiration there, as scientists continue to develop new understanding of brain function.

Theory and Examples

During the late 1960s and the 1970s the number of researchers in the field of neural networks dropped dramatically. There were, however, a number of researchers who continued to work during this period, including Tuevo Kohonen, James Anderson, Kunihiko Fukushima and Shun-ichi Amari, among others. One of the most prolific was Stephen Grossberg.

Grossberg has been continuously active, and highly productive, in neural network research since the early 1960s. His work is characterized by the use of nonlinear mathematics to model specific functions of mind and brain, and his volume of output has been consistent with the magnitude of the task of understanding the brain. The topics of his papers have ranged from such specific areas as how competitive networks can provide contrast enhancement in vision, to such general subjects as a universal theory for human memory.

In part because of the scale of his efforts, his work has a reputation for difficulty. Each new paper is built on a foundation of 30 years of previous results, and is therefore difficult to assess on its own merits. In addition, his terminology is self-consistent, but not in standard use by other researchers. His work is also characterized by a high level of mathematical and neurophysiological sophistication. He is inspired by the interdisciplinary research into brain function by Helmholtz, Maxwell and Mach, and he brings this viewpoint to his work. His research lies at the intersection of mathematics, psychology and neurophysiology. A lack of background in these areas can make his work difficult to approach on a first reading.

This chapter will take a rudimentary look at one of the seminal networks developed by Grossberg. In order to obtain the maximum understanding of his ideas, we will begin with a brief introduction to the biological motivation for the network: the visual system. Then we will present the mathematical building block for many of Grossberg's networks: the shunting model. After understanding the function of this simple model, we will demonstrate how it can be used to build a neural network for adaptive pattern recognition. This network will then form the basis for the adaptive resonance theory networks that are discussed in Chapter 19. By building up gradually to the more complex networks, we hope to make them more easily understandable.

There is an important lesson we should take from the work described in this chapter. Although the original inspiration for the field of artificial neural networks came from biology, at times we forget to look back to biology for new ideas. It will be the blending of biology, mathematics, psychology and other disciplines that will provide the maximum growth in our understanding of neural networks.

Biological Motivation: Vision

The neural network described in this chapter was inspired by the developmental physiology of the human visual system. In this section we want to provide a brief introduction to vision, so that the function of the network will be more understandable.

In Figure 18.1 we have a schematic representation of the first stages of the visual system. Light passes through the cornea (the transparent front part of the eye) and the lens, which bends the light to focus objects on the retina (the interior layer of the external wall of the eye). It is after the light falls on the retina that the immense job of translating this information into an understandable image begins. As we will see later in this chapter, much of what we "see" is not actually present in the image projected onto the retina.

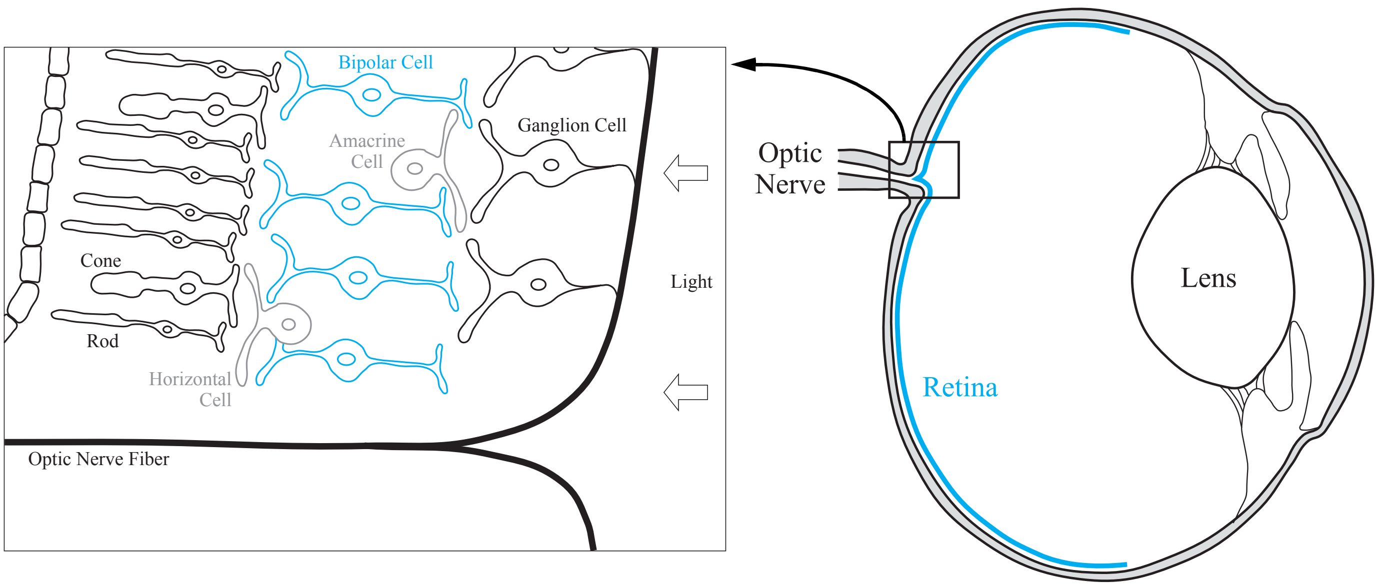


Figure 18.1 Eyeball and Retina

Retina

The *retina* is actually a part of the brain. It becomes separated from the brain during fetal development, but remains connected to it through the optic nerve. The retina consists of three layers of nerve cells. The outer layer consists of the photoreceptors (rods and cones), which convert light into electrical signals. The *rods* allow us to see in dim light, whereas the *cones* allow us to see fine detail and color. For reasons not completely understood, light must pass through the other two layers of the retina in order to stimulate the rods and cones. As we will see later, this obstruction must be compensated for in neural processing, in order to reconstruct recognizable images.

Rods

Cones

Bipolar Cells

The middle layer of the retina consists of three types of cells: bipolar cells, horizontal cells and amacrine cells. *Bipolar cells* receive input from the receptors and feed into the third layer of the retina, containing the ganglion

Horizontal Cells
Amacrine Cells

cells. *Horizontal cells* link the receptors and the bipolar cells, and *amacrine cells* link bipolar cells with the ganglion cells.

Ganglion Cells

The final layer of the retina contains the *ganglion cells*. The axons of the ganglion cells pass across the surface of the retina and collect in a bundle to form the optic nerve. It is interesting to note that each eye contains roughly 125 million receptors, but only 1 million ganglion cells. Clearly there is significant processing done in the retina to perform data reduction.

Visual Cortex

The axons of the ganglion cells, bundled into the optic nerve, connect to an area of the brain called the lateral geniculate nucleus, as illustrated in Figure 18.2. From this point the fibers fan out into the primary visual cortex, located at the back of the brain. The axons of the ganglion cells make synapses with lateral geniculate cells, and the axons of the lateral geniculate cells make synapses with cells in the visual cortex. The *visual cortex* is the region of the brain devoted to visual function and consists of many layers of cells.

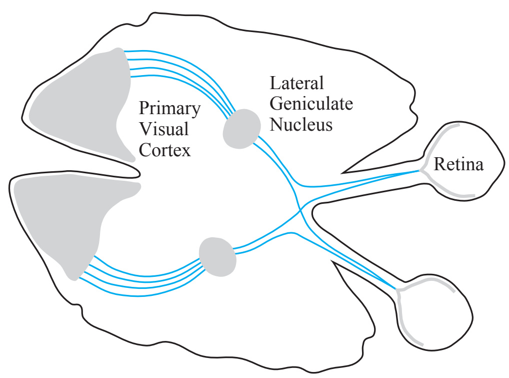


Figure 18.2 Visual Pathway

The connections along the visual pathway are far from random. The mapping from each layer to the next is highly organized. The axons from the ganglion cells in a certain part of the retina go to cells in a particular part of the lateral geniculate, which in turn go to a particular part of the visual cortex. (This topographic mapping was one of the inspirations for the self-organizing feature map described in Chapter 14.) In addition, as we can see in Figure 18.2, each hemisphere of the brain receives input from both eyes, since half of the optic nerve fibers cross and the other half stay uncrossed. It turns out that the left half of each visual field ends up in the right half of the brain, and the right half of each visual field ends up in the left half of the brain.

Illusions

We now have some idea of the general structure of the visual pathway, but how does it function? What is the purpose of the three layers of the retina? What operations are performed by the lateral geniculate? Some hints to the answers to these questions can be obtained by investigating visual illusions.

Why are there so many visual illusions? Mechanisms that overcome imperfections of the retinal uptake process imply illusions. Grossberg and others have used the vast store of known illusions to probe adaptive perceptual mechanisms [GrMi89]. If we can develop mathematical models that produce the same illusions the biological system does, then we may have a mechanism that describes how this part of the brain works. To help us understand why illusions exist, we will first consider some of the imperfections of the retinal uptake process.

Optic Disk

Figure 18.3 is the view of the retina that an ophthalmologist has when looking into the eye through the cornea. The large pale circle is the *optic disk*, where the optic nerve leaves the retina on its way to the lateral geniculate. This is also where arteries enter and veins leave the retina. The optic disk causes a blind spot in our vision, as we will discuss in a moment.

Fovea

The darker disk to the right of the optic disk is the *fovea*, which constitutes the center of our field of vision. This is a region of the retina, about half a millimeter in diameter, that contains only cones. Although cones are distributed throughout the retina, they are most densely packed in the fovea. In addition, in this area of the retina the other layers are displaced to the side, so that the cones lie at the front. The densely packed photoreceptors, and the lack of obstruction, give us our best fine-detail vision at the fovea, which allows us to precisely focus the lens.

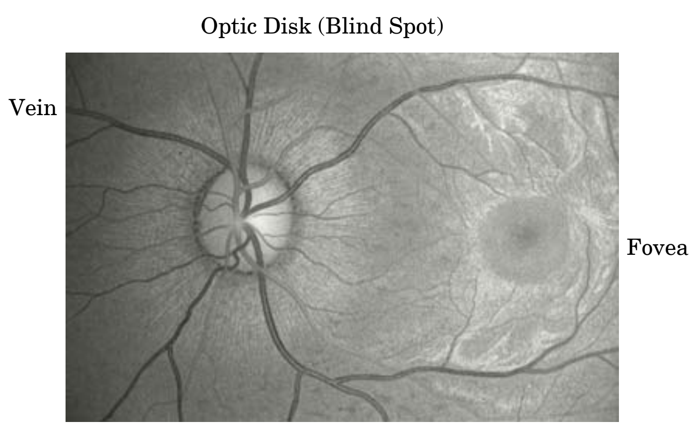


Figure 18.3 Back of the Eye (from [John01])

From Figure 18.3 we can see that there are a number of imperfections in retinal uptake. First, there are no rods and cones in the optic disk, which leaves a blind spot in our field of vision. We are not normally aware of the blind spot because of processing done in the visual pathway, but we can identify it with a simple test. Look at the blue circle on the left side of Figure 18.4, while covering your left eye. As you move your head closer to the page, then farther away, you will notice a point (about nine inches away) at which the circle on the right will disappear from your field of vision. (You are still looking at the circle on the left.) If you haven't tried this before, it can be a little disconcerting. The interesting thing is that we don't see our blind spot as a black hole. Somehow our brains fill in the missing region.

Figure 18.4 Test for the Blind Spot

Other imperfections in the retinal uptake are the arteries and veins that cross in front of the photoreceptors at the back of the retina. These obstruct the rods and cones from receiving all of the light in the visual field. In addition, because the photoreceptors are at the back of the retina, light must pass through the other two layers to reach them.

Figure 18.5 illustrates the effect of these imperfections. Here we see an edge displayed on the retina. The drawing on the right illustrates the image initially perceived by the photoreceptors. The regions covered by the blind spot and the veins are not observed by the rods and cones. (The reason we do not "see" the arteries, veins, etc., is that the vision pathway does not respond to stabilized images. The eyeballs are constantly jerking, in what are called saccadic movements, so that even fixed objects in our field of vision are moving relative to the eye. The veins are fixed relative to the eyeball, so they fade from our vision.)

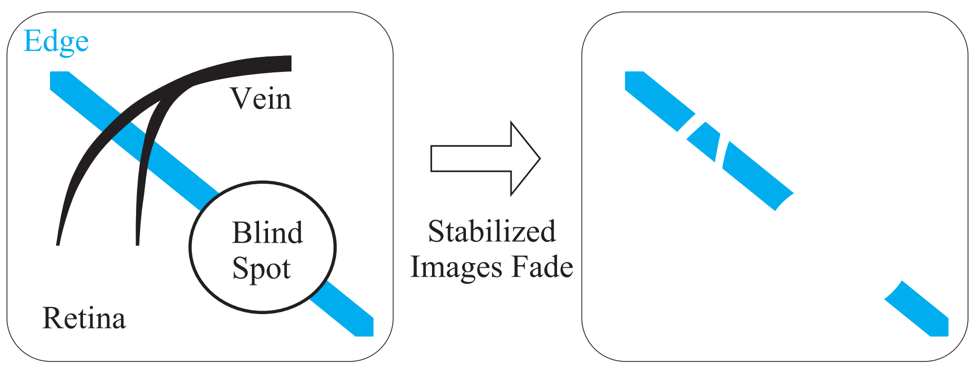


Figure 18.5 Perception of an Edge on the Retina (after [Gros90])

Emergent Segmentation
Featural Filling-in

Because we do not see edges as displayed on the right side of Figure 18.5, the neural systems in our visual pathway must be performing some operation that compensates for the distortions and completes the image. Grossberg suggests [GrMi89] that there are two primary types of compensatory processing involved. The first, which he calls *emergent segmentation*, completes missing boundaries. The second, which he calls *featural filling-in*, fills in the color and brightness inside the resulting boundaries. These two processes are illustrated in Figure 18.6. In the top figure we see an edge as it is originally perceived by the rods and cones, with missing sections. In the lower figure we see the completed edge, after the emergent segmentation and featural filling-in.

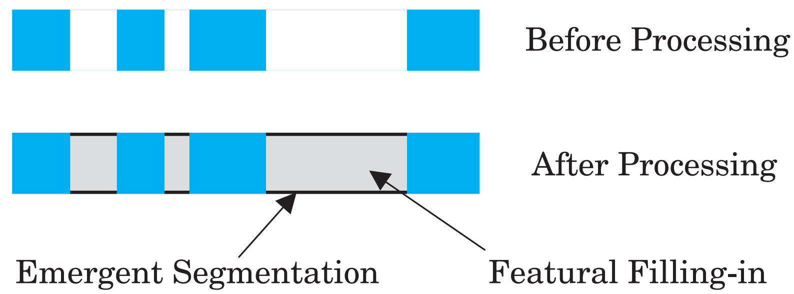
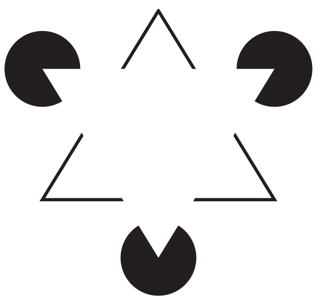
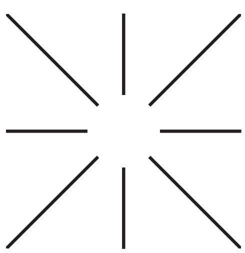


Figure 18.6 Compensatory Processing (after [Gros90])





If the processing along the visual pathway is recreating missing parts of the images we see, there must be times when it makes mistakes, since it cannot know exactly those parts of a scene from which it receives no light. These mistakes are illustrated by visual illusions. Consider, for example, the two figures in the left margin. In the top figure you should be able to see a bright white triangle lying on top of several other black objects. In fact, no such triangle exists in the figure. It is purely a creation of the emergent segmentation and featural filling-in process of your visual system. The same is true of the bright white circle which appears to lie on top of the lines in the lower-left figure.

The featural filling-in process is also demonstrated in Figure 18.7. This illusion is called neon color spreading [vanT75]. In the diagram on the right you may be able to see light blue diamonds, or even wide light blue lines criss-crossing the figure. In the diagram on the left you may be able to see a light blue ring. The blue you see filling in the diamonds and the ring is not a result of the color having been smeared during the printing process, nor is it caused by the scattering of light. This effect does not appear on the retina at all. It does not exist, except in your brain. (The perception of neon color spreading can vary from individual to individual, and the strength of the perception is dependent on the colors used. If you do not notice the effect in Figure 18.7, look at the cover of any issue of the journal *Neural Networks*, Pergamon Press.)

Later in this chapter we will discuss some neural network models that can help to explain the processes that implement emergent segmentation, as well as other visual phenomena.

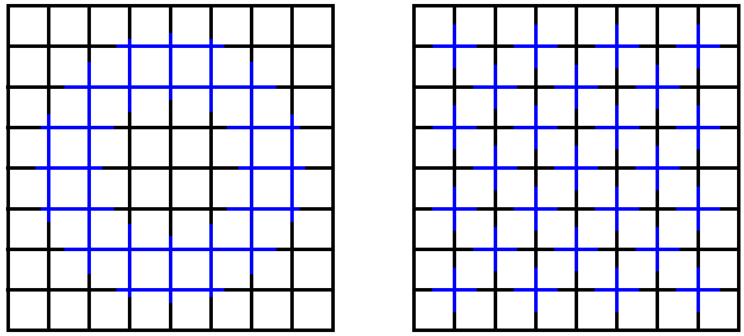


Figure 18.7 Neon Color Spreading (Featural Filling In)

Vision Normalization

Brightness Constancy
Brightness Contrast

In addition to emergent segmentation and featural filling-in, there are two other phenomena that give us an indication of what operations are being performed in the early vision system: *brightness constancy* and *brightness contrast*. The brightness constancy effect is evidenced by the test illustrated in Figure 18.8. In this test a subject is shown a small grey disk inside a darker grey annulus, which is illuminated by white light of a certain intensity. The subject is asked to indicate the brightness of the central disk by looking at a series of grey disks, separately illuminated, and selecting the disk with the same brightness. Next, the brightness of the light illuminating the grey disk and dark annulus is increased, and the subject is again asked to select the disk with the same brightness. This process is repeated for several different levels of illumination. It turns out that in each case the subject will choose the same disk as matching the original central disk. Even though the total light entering the subject's eye is 10 to 100 times brighter, it is only the relative brightness that registers.

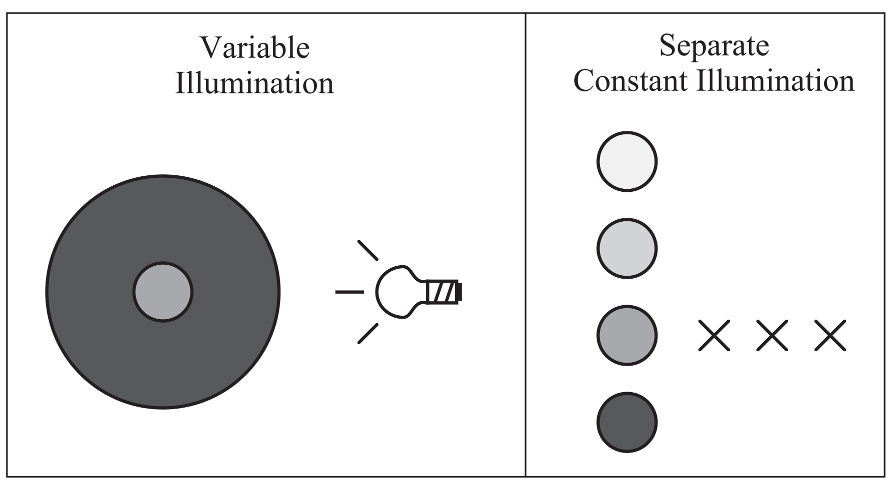
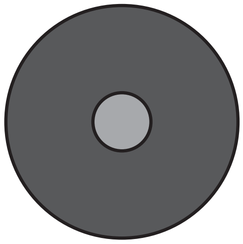
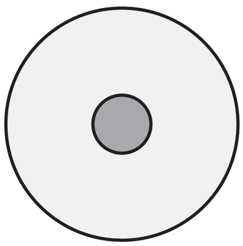


Figure 18.8 Test of Brightness Constancy (after [Gros90])





Another phenomenon of the vision system, which is closely related to brightness constancy, is brightness contrast. This effect is illustrated by the two figures in the left margin. At the centers of the two figures we have two small disks with equivalent grey scale. The small disk in the top figure is surrounded by a darker annulus, while the small disk in the lower figure is surrounded by a lighter annulus. Even though both disks have the same grey scale, the one inside the darker annulus appears brighter. This is because our vision system is sensitive to relative intensities. It would seem that the total activity across the image is held constant.

The properties of brightness constancy and brightness contrast are very important to our vision system. Since we see things in so many different lighting conditions, if we were not able to compensate for the absolute intensity of a scene, we would never learn to recognize things. Grossberg calls this process of normalization "discounting the illuminant."

In the rest of this chapter we want to present a neural network architecture that is consistent with the physical phenomena discussed in this section.

Basic Nonlinear Model

Before we introduce the Grossberg network, we will begin by looking at some of the building blocks that make up the network. The first building block is the *"leaky" integrator*, which is shown in Figure 18.9. The basic equation for this system is

Leaky Integrator

$$\varepsilon \frac{dn(t)}{dt} = -n(t) + p(t) , \tag{18.1}$$

Time Constant

where ε is the system *time constant*.

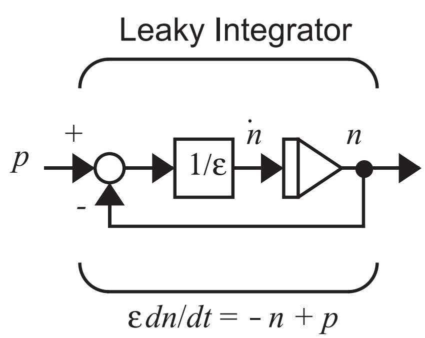


Figure 18.9 Leaky Integrator

The response of the leaky integrator to an arbitrary input $p(t)$ is

$$n(t) = e^{-t/\varepsilon} n(0) + \frac{1}{\varepsilon} \int_0^t e^{-(t-\tau)/\varepsilon} p(\tau) d\tau . \tag{18.2}$$

For example, if the input $p(t)$ is constant and the initial condition $n(0)$ is zero, Eq. (18.2) will produce

$$n(t) = p(1 - e^{-t/\varepsilon}) . \tag{18.3}$$

A graph of this response, for $p = 1$ and $\varepsilon = 1$, is given in Figure 18.10. The response exponentially approaches a steady state value of 1.

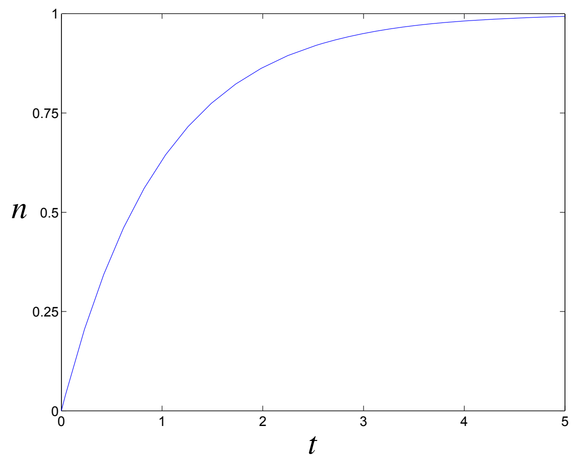


Figure 18.10 Leaky Integrator Response

There are two important properties of the leaky integrator that we want to note. First, because Eq. (18.1) is linear, if the input p is scaled, then the response $n(t)$ will be scaled by the same amount. For example, if the input is doubled, then the response will also be doubled, but will maintain the same shape. This is evident in Eq. (18.3). Second, the speed of response of the leaky integrator is determined by the time constant ε. When ε decreases, the response becomes faster; when ε increases, the response becomes slower. (See Problem P18.1.)

To experiment with the leaky integrator, use the Neural Network Design Demonstration Leaky Integrator (`nnd15li`).

Shunting Model

The leaky integrator forms the nucleus of one of Grossberg's fundamental neural models: the *shunting model*, which is shown in Figure 18.11. The equation of operation of this network is

$$\varepsilon \frac{dn(t)}{dt} = -n(t) + (b^{+} - n(t))p^{+} - (n(t) + b^{-})p^{-} , \tag{18.4}$$

Excitatory

Inhibitory

where p^{+} is a nonnegative value representing the *excitatory* input to the network (the input that causes the response to increase), and p^{-} is a nonnegative value representing the *inhibitory* input (the input that causes the response to decrease). The biases b^{+} and b^{-} are nonnegative constants that determine the upper and lower limits on the neuron response, as we will explain next.

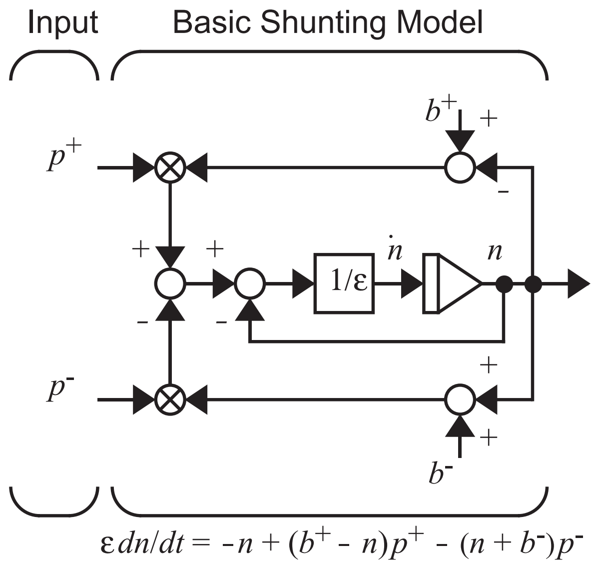
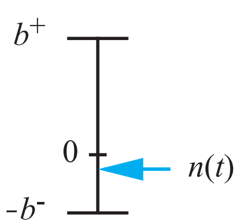


Figure 18.11 Shunting Model

There are three terms on the right-hand side of Eq. (18.4). When the net sign of these three terms is positive, $n(t)$ will go up. When the net sign is negative, $n(t)$ will go down. To understand the performance of the network, let's investigate the three terms individually.

The first term, $-n(t)$, is a linear decay term, which is also found in the leaky integrator. It is negative whenever $n(t)$ is positive, and positive whenever $n(t)$ is negative. The second term, $(b^+ - n(t))p^+$, provides nonlinear gain control. This term will be positive while $n(t)$ is less than b^+, but will become zero when $n(t) = b^+$. This effectively sets an upper limit on $n(t)$ of b^+. The third term, $-(n(t) + b^-)p^-$, also provides nonlinear gain control. It sets a lower limit on $n(t)$ of $-b^-$.

Figure 18.12 illustrates the performance of the shunting network when $b^+ = 1$, $b^- = 0$ and $\varepsilon = 1$. In the left graph we see the network response when the excitatory input $p^+ = 1$ and the inhibitory input $p^- = 0$. For the right graph $p^+ = 5$ and $p^- = 0$. Notice that even though the excitatory input is increased by a factor of 5, the steady state network response is increased by less than a factor of 2. If we were to continue to increase the excitatory input, we would find that the steady state network response would increase, but would always be less than $b^+ = 1$.



If we apply an inhibitory input to the shunting network, the steady state network response will decrease, but will remain greater than $-b^-$. To summarize the operation of the shunting model, if $n(0)$ falls between b^+ and $-b^-$, then $n(t)$ will remain between these limits, as shown in the figure in the left margin.

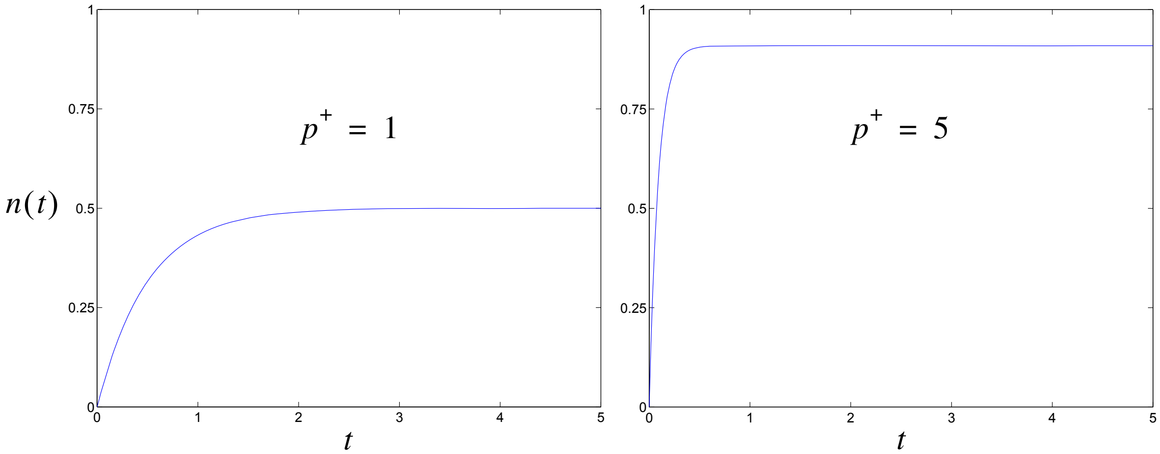


Figure 18.12 Shunting Network Response

The shunting model will form the basis for the Grossberg competitive network, which we will discuss in the next section. The nonlinear gain control will be used to normalize input patterns and to maintain relative intensities over a wide range of total intensity.

To experiment with the shunting network, use the Neural Network Design Demonstration Shunting Network (`nnd15sn`).

Two-Layer Competitive Network

We are now ready to present the Grossberg competitive network. This network was inspired by the operation of the mammalian visual system, which we discussed in the opening section of this chapter. (Grossberg was influenced by the work of Chistoph von der Malsburg [vond73], which was influenced in turn by the Nobel-prize-winning experimental work of David Hubel and Torsten Wiesel [HuWi62].) A block diagram of the network is shown in Figure 18.13.

Short-Term Memory
Long-Term Memory

There are three components to the Grossberg network: Layer 1, Layer 2 and the adaptive weights. Layer 1 is a rough model of the operation of the retina, while Layer 2 represents the visual cortex. These models do not fully explain the complexity of the human visual system, but they do illustrate a number of its characteristics. The network includes *short-term memory* (STM) and *long-term memory* (LTM) mechanisms, and performs adaptation, filtering, normalization and contrast enhancement. In the following subsections we will discuss the operation of each of the components of the network.

As we analyze the various elements of the Grossberg network, you will notice the similarity to the Kohonen competitive network of the previous chapter.

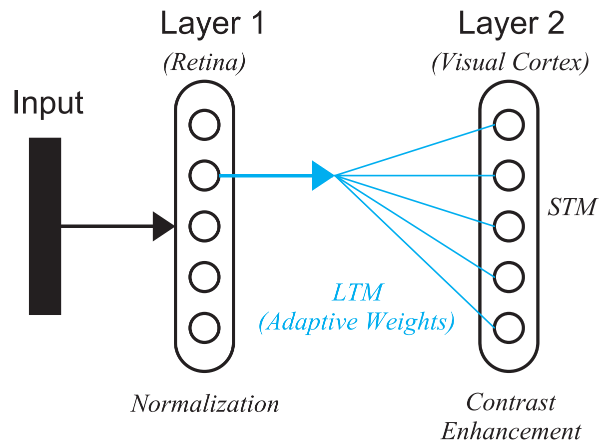


Figure 18.13 Grossberg Competitive Network

Layer 1

Layer 1 of the Grossberg network receives external inputs and normalizes the intensity of the input pattern. (Recall from Chapter 14 that the Kohonen network performs best when the input patterns are normalized. For the Grossberg network the normalization is accomplished by the first layer of the network.) A block diagram of this layer is given in Figure 18.14. Note that it uses the shunting model, with the excitatory and inhibitory inputs computed from the input vector $\mathbf{p}$.

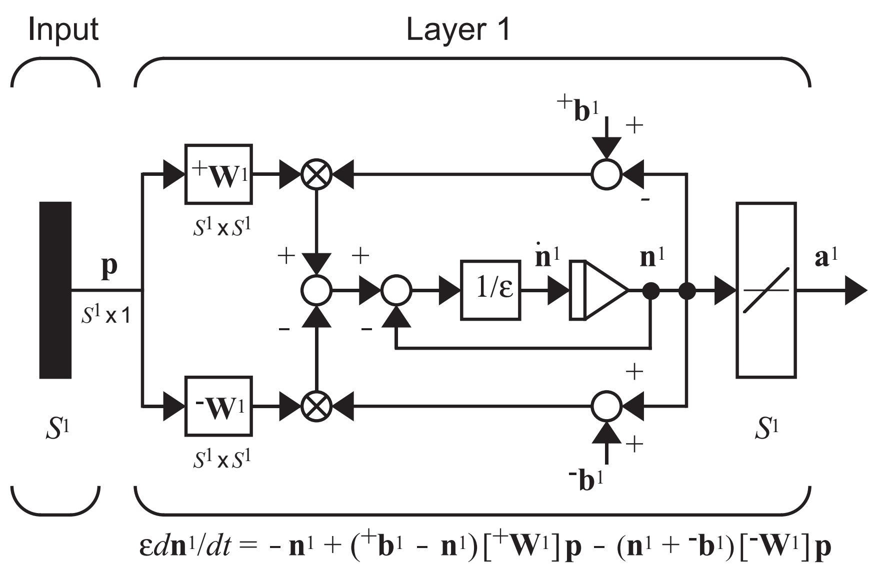


Figure 18.14 Layer 1 of the Grossberg Network

The equation of operation of Layer 1 is

$$\varepsilon\frac{d\mathbf{n}^1(t)}{dt} = -\mathbf{n}^1(t) + ({}^{+}\mathbf{b}^1 - \mathbf{n}^1(t))[{}^{+}\mathbf{W}^1]\mathbf{p} - (\mathbf{n}^1(t) + {}^{-}\mathbf{b}^1)[{}^{-}\mathbf{W}^1]\mathbf{p} . \qquad (18.5)$$

As we mentioned earlier, the parameter ε determines the speed of response. It is chosen so that the neuron responses will be much faster than the changes in the adaptive weights, which we will discuss in a later section.

Eq. (18.5) is a shunting model with excitatory input $[{}^{+}\mathbf{W}^{1}]\mathbf{p}$, where

$$ {}^{+}\mathbf{W}^{1} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} . \qquad (18.6) $$

Therefore the excitatory input to neuron i is the ith element of the input vector.

The inhibitory input to Layer 1 is $[{}^{-}\mathbf{W}^{1}]\mathbf{p}$, where

$$ {}^{-}\mathbf{W}^{1} = \begin{bmatrix} 0 & 1 & \cdots & 1 \\ 1 & 0 & \cdots & 1 \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \cdots & 0 \end{bmatrix} . \qquad (18.7) $$

Thus the inhibitory input to neuron i is the sum of all elements of the input vector, except the ith element.

On-Center/Off-Surround

The connection pattern defined by the matrices ${}^{+}\mathbf{W}^{1}$ and ${}^{-}\mathbf{W}^{1}$ is called an *on-center/off-surround* pattern. This is because the excitatory input for neuron i (which turns the neuron on) comes from the element of the input vector centered at the same location (element i), while the inhibitory input (which turns the neuron off) comes from surrounding locations. This type of connection pattern produces a normalization of the input pattern, as we will show in the following discussion.

For simplicity, we will set the inhibitory bias ${}^{-}\mathbf{b}^{1}$ to zero, which sets the lower limit of the shunting model to zero, and we will set all elements of the excitatory bias ${}^{+}\mathbf{b}^{1}$ to the same value, i.e.,

$$ {}^{+}b_i^1 = {}^{+}b^1 ,\ i = 1, 2, \dots , S^1 , \qquad (18.8) $$

so that the upper limit for all neurons will be the same.

To investigate the normalization effect of Layer 1, consider the response of neuron i:

$$\varepsilon \frac{dn_i^1(t)}{dt} = -n_i^1(t) + ({}^+b^1 - n_i^1(t))p_i - n_i^1(t) \sum_{j \neq i} p_j . \tag{18.9}$$

In the steady state ($dn_i^1(t)/dt = 0$) we have

$$0 = -n_i^1 + ({}^+b^1 - n_i^1)p_i - n_i^1 \sum_{j \neq i} p_j . \tag{18.10}$$

If we solve for the steady state neuron output n_i^1 we find

$$n_i^1 = \frac{{}^+b^1 p_i}{1 + \sum_{j=1}^{S^1} p_j} . \tag{18.11}$$

We now define the relative intensity of input i to be

$$\bar{p}_i = \frac{p_i}{P} \text{ where } P = \sum_{j=1}^{S^1} p_j . \tag{18.12}$$

Then the steady state neuron activity can be written

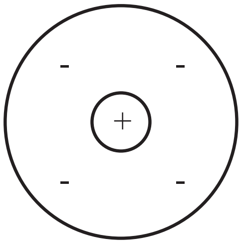
$$n_i^1 = \left(\frac{{}^+b^1 P}{1 + P}\right) \bar{p}_i . \tag{18.13}$$

Therefore n_i^1 will be proportional to the relative intensity $\bar{p}_i$, regardless of the magnitude of the total input P. In addition, the total neuron activity is bounded:

$$\sum_{j=1}^{S^1} n_j^1 = \sum_{j=1}^{S^1} \left(\frac{{}^+b^1 P}{1 + P}\right) \bar{p}_j = \left(\frac{{}^+b^1 P}{1 + P}\right) \leq {}^+b^1 . \tag{18.14}$$

The input vector is normalized so that the total activity is less than ${}^+b^1$, while the relative intensities of the individual elements of the input vector are maintained. Therefore, the outputs of Layer 1, n_i^1, code the relative input intensities, $\bar{p}_i$, rather than the instantaneous fluctuations in the total input activity, P. This result is produced by the on-center/off-surround connection pattern of the inputs and the nonlinear gain control of the shunting model.

Note that Layer 1 of the Grossberg network explains the brightness constancy and brightness contrast characteristics of the human visual system, which we discussed on page 15-8. The network is sensitive to the relative



intensities of an image, rather than absolute intensities. In addition, experimental evidence has shown that the on-center/off-surround connection pattern is a characteristic feature of the receptive fields of retinal ganglion cells [Hube88]. (The receptive field is an area of the retina in which the photoreceptors feed into a given cell. The figure in the left margin illustrates the on-center/off-surround receptive field of a typical retinal ganglion cell. A "+" indicates an excitatory region, and a "-" indicates an inhibitory region. It is a two-dimensional pattern, as opposed to the one-dimensional connections of Eq. (18.6) and Eq. (18.7).)

To illustrate the performance of Layer 1, consider the case of two neurons, with ${}^{+}b^1 = 1$, $\varepsilon = 0.1$:

$$(0.1)\frac{dn_1^1(t)}{dt} = -n_1^1(t) + (1 - n_1^1(t))p_1 - n_1^1(t)p_2 , \tag{18.15}$$

$$(0.1)\frac{dn_2^1(t)}{dt} = -n_2^1(t) + (1 - n_2^1(t))p_2 - n_2^1(t)p_1 . \tag{18.16}$$

The response of this network, for two different input vectors, is shown in Figure 18.15. For both input vectors, the second element is four times as large as the first element, although the total intensity of the second input vector is five times as large as that of the first input vector (50 vs. 10). From Figure 18.15 we can see that the response of the network maintains the relative intensities of the inputs, while limiting the total response. The total response ($n_1^1(t) + n_2^1(t)$) will always be less than 1.

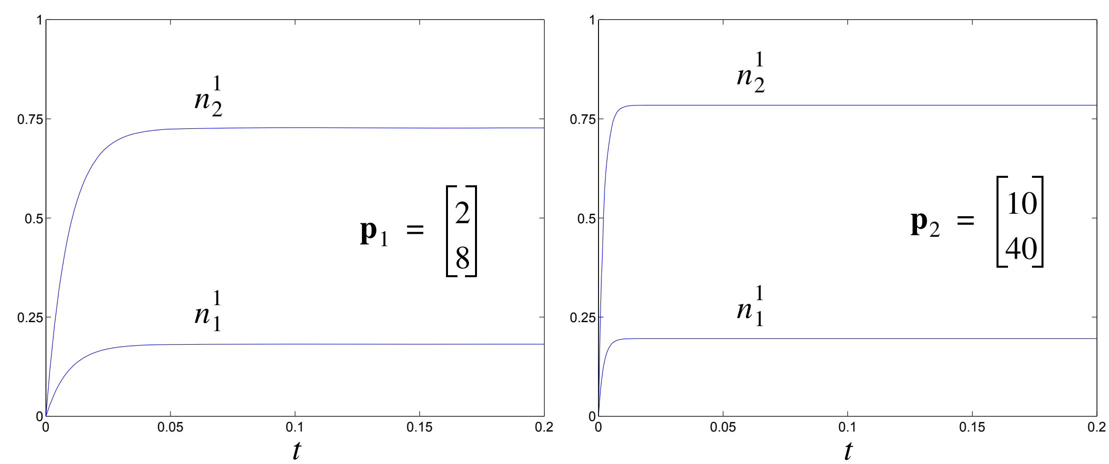


Figure 18.15 Layer 1 Response

To experiment with Layer 1 of the Grossberg network, use the Neural Network Design Demonstration Grossberg Layer 1 (`nnd15gl1`).

Layer 2

Layer 2 of the Grossberg network, which is a layer of continuous-time instars, performs several functions. First, like Layer 1, it normalizes total activity in the layer. Second, it contrast enhances its pattern, so that the neuron that receives the largest input will dominate the response. (This is closely related to the winner-take-all competition in the Hamming network and the Kohonen network.) Finally, it operates as a *short-term memory* (STM) by storing the contrast-enhanced pattern.

Short-Term Memory

Figure 18.16 is a diagram of Layer 2. As with Layer 1, the shunting model forms the basis for Layer 2. The main difference between Layer 2 and Layer 1 is that Layer 2 uses feedback connections. The feedback enables the network to store a pattern, even after the input has been removed. The feedback also performs the competition that causes the contrast enhancement of the pattern. We will demonstrate these properties in the following discussion.

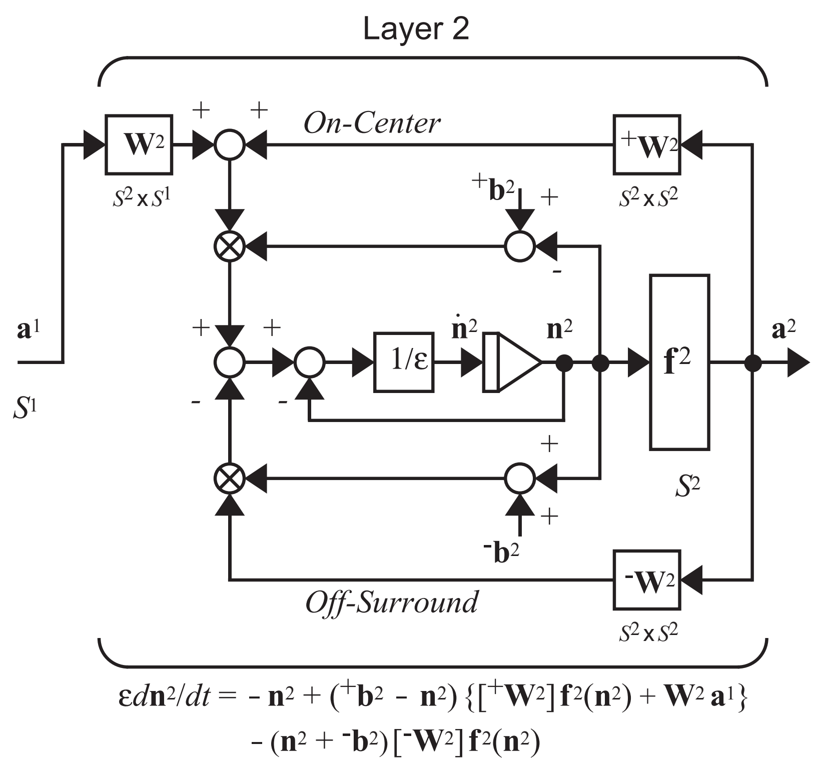


Figure 18.16 Layer 2 of the Grossberg Network

The equation of operation of Layer 2 is

$$\varepsilon\frac{d\mathbf{n}^2(t)}{dt} = -\mathbf{n}^2(t) + ({}^{+}\mathbf{b}^2 - \mathbf{n}^2(t))\{[{}^{+}\mathbf{W}^2]\mathbf{f}^2(\mathbf{n}^2(t)) + \mathbf{W}^2\mathbf{a}^1\}$$
$$-(\mathbf{n}^2(t) + {}^{-}\mathbf{b}^2)[{}^{-}\mathbf{W}^2]\mathbf{f}^2(\mathbf{n}^2(t)) \qquad (18.17)$$

This is a shunting model with excitatory input $\{[{}^{+}\mathbf{W}^2]\mathbf{f}^2(\mathbf{n}^2(t)) + \mathbf{W}^2\mathbf{a}^1\}$, where ${}^{+}\mathbf{W}^2 = {}^{+}\mathbf{W}^1$ provides on-center feedback connections, and $\mathbf{W}^2$ consists of adaptive weights, analogous to the weights in the Kohonen network. The rows of $\mathbf{W}^2$, after training, will represent the prototype patterns. The inhibitory input to the shunting model is $[{}^{-}\mathbf{W}^2]\mathbf{f}^2(\mathbf{n}^2(t))$, where ${}^{-}\mathbf{W}^2 = {}^{-}\mathbf{W}^1$ provides off-surround feedback connections.

To illustrate the performance of Layer 2, consider a two-neuron layer with

$$\varepsilon = 0.1\,,\ {}^{+}\mathbf{b}^2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix},\ {}^{-}\mathbf{b}^2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix},\ \mathbf{W}^2 = \begin{bmatrix} ({}_1\mathbf{w}^2)^T \\ ({}_2\mathbf{w}^2)^T \end{bmatrix} = \begin{bmatrix} 0.9 & 0.45 \\ 0.45 & 0.9 \end{bmatrix}, \qquad (18.18)$$

and

$$f^2(n) = \frac{10(n)^2}{1+(n)^2}\,. \qquad (18.19)$$

The equations of operation of the layer will be

$$(0.1)\frac{dn_1^2(t)}{dt} = -n_1^2(t) + (1 - n_1^2(t))\{f^2(n_1^2(t)) + ({}_1\mathbf{w}^2)^T\mathbf{a}^1\} - n_1^2(t)f^2(n_2^2(t)) \qquad (18.20)$$

$$(0.1)\frac{dn_2^2(t)}{dt} = -n_2^2(t) + (1 - n_2^2(t))\{f^2(n_2^2(t)) + ({}_2\mathbf{w}^2)^T\mathbf{a}^1\} - n_2^2(t)f^2(n_1^2(t))\,. \qquad (18.21)$$

Note the relationship between these equations and the Hamming and Kohonen networks. The inputs to Layer 2 are the inner products between the prototype patterns (rows of the weight matrix $\mathbf{W}^2$) and the output of Layer 1 (normalized input pattern). The largest inner product will correspond to the prototype pattern closest to the input pattern. Layer 2 then performs a competition between the neurons, which tends to *contrast enhance* the output pattern — maintaining large outputs while attenuating small outputs. This contrast enhancement is generally milder than the winner-take-all competition of the Hamming and Kohonen networks. In the Hamming and Kohonen networks, the competition drives all but one of the neuron outputs to zero. The exception is the one with the largest input. In the Grossberg network, the competition maintains large values and attenuates small values, but does not necessarily drive all small values to zero. The amount

Contrast Enhance

of contrast enhancement is determined by the transfer function f^2, as we will see in the next section.

Figure 18.17 illustrates the response of Layer 2 when the input vector $\mathbf{a}^1 = \begin{bmatrix} 0.2 & 0.8 \end{bmatrix}^T$ (the steady state result obtained from our Layer 1 example) is applied for 0.25 seconds and then removed.

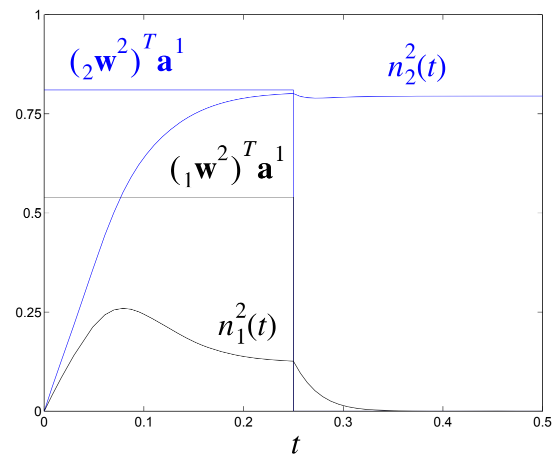


Figure 18.17 Layer 2 Response

There are two important characteristics of this response. First, even before the input is removed, some contrast enhancement is performed. The inputs to Layer 2 are

$$({}_1\mathbf{w}^2)^T\mathbf{a}^1 = \begin{bmatrix} 0.9 & 0.45 \end{bmatrix} \begin{bmatrix} 0.2 \\ 0.8 \end{bmatrix} = 0.54 , \qquad (18.22)$$

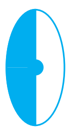
$$({}_2\mathbf{w}^2)^T\mathbf{a}^1 = \begin{bmatrix} 0.45 & 0.9 \end{bmatrix} \begin{bmatrix} 0.2 \\ 0.8 \end{bmatrix} = 0.81 . \qquad (18.23)$$

Therefore the second neuron has 1.5 times as much input as the first neuron. However, after 0.25 seconds the output of the second neuron is 6.34 times the output of the first neuron. The contrast between high and low has been increased dramatically.

The second characteristic of the response is that after the input has been set to zero, the network further enhances the contrast and stores the pattern. In Figure 18.17 we can see that after the input is removed (at 0.25 seconds) the output of the first neuron decays to zero, while the output of the second neuron reaches a steady state value of 0.79. This output is maintained, even after the input is removed. (Grossberg calls this behavior *reverberation* [Gross76].) It is the nonlinear feedback that enables the network to store the pattern, and the on-center/off-surround connection pattern (determined by ${}^{+}\mathbf{W}^{2}$ and ${}^{-}\mathbf{W}^{2}$) that causes the contrast enhancement.

As an aside, note that we have used the on-center/off-surround structure in both layers of the Grossberg network. There are other connection patterns that could be used for different applications. Recall, for instance, the emergent segmentation problem discussed earlier in this chapter. A structure that has been proposed to implement this mechanism is the *oriented receptive field* [GrMi89], which is shown in the left margin. For this structure, the "on" (excitatory) connections come from one side of the field (indicated by the blue region), and the "off" (inhibitory) connections come from the other side of the field (indicated by the white region).

Oriented Receptive Field



The operation of the oriented receptive field is illustrated in Figure 18.18. When the field is aligned with an edge, the corresponding neuron is activated (large response). When the field is not aligned with an edge, then the neuron is inactive (small response). This explains why we might perceive an edge where none exists, as can be seen in the right-most receptive field shown in Figure 18.18.

For a complete discussion of oriented receptive fields and how they can be incorporated into a neural network architecture for preattentive vision, see [GrMi89]. This paper also discusses a mechanism for featural filling-in.

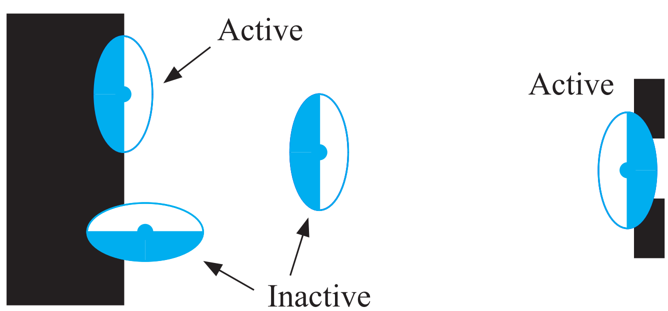
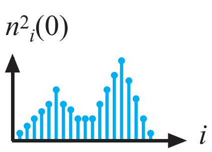


Figure 18.18 Operation of Oriented Receptive Field

Choice of Transfer Function



The behavior of Layer 2 of the Grossberg network depends very much on the transfer function $f^2(n)$. For example, suppose that an input has been applied for some length of time, so that the output has stabilized to the pattern shown in the left margin. (Each point represents the output of an individual neuron.) If the input is then removed, Figure 18.19 demonstrates how the choice of $f^2(n)$ will affect the steady state response of the network. (See [Gross82].)

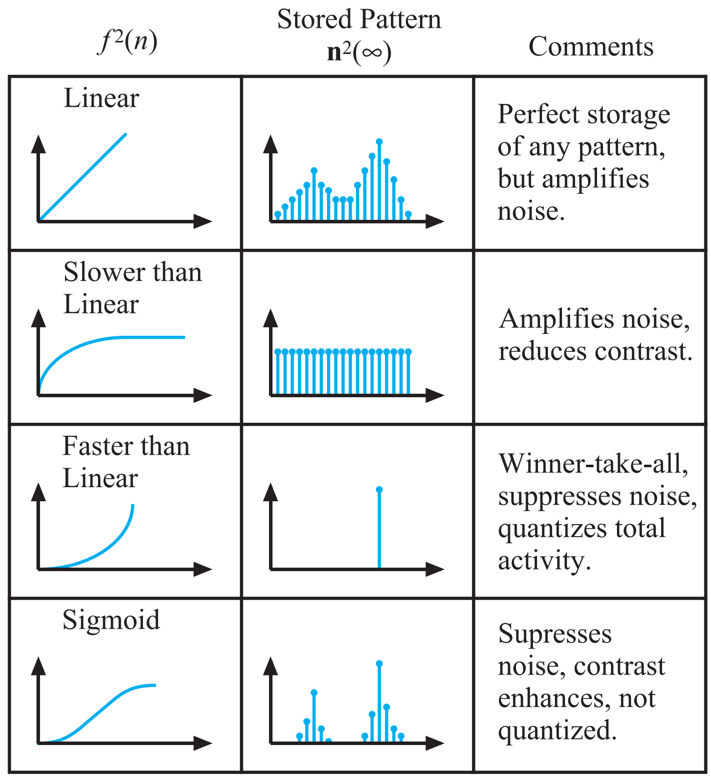


Figure 18.19 Effect of Transfer Function $f^2(n)$ (after [Gross82])

If the transfer function is *linear*, the pattern is perfectly stored. Unfortunately, the noise in the pattern will be amplified and stored as easily as the significant inputs. (See Problem P18.6.) If the transfer function is *slower-than-linear* (e.g., $f^2(n) = 1 - e^{-n}$), the steady state response is independent of the initial conditions; all neurons that begin with nonzero values will come to the same level in the steady state. All contrast is eliminated, and noise is amplified.

Faster-than-linear transfer functions (e.g., $f^2(n) = (n)^2$) produce a winner-take-all competition. Only those neurons with the largest initial values are stored; all others are driven to zero. This minimizes the effect of noise, but quantizes the response into an all-or-nothing signal (as in the Hamming and Kohonen networks).

A *sigmoid* function is faster-than-linear for small signals, approximately linear for intermediate signals and slower-than-linear for large signals. When a sigmoid transfer function is used in Layer 2, the pattern is contrast enhanced; larger values are amplified, and smaller values are attenuated. All initial neuron outputs that are less than a certain level (called the *quenching threshold* by Grossberg [Gros76]) decay to zero. This merges the noise suppression of the faster-than-linear transfer functions with the perfect storage produced by linear transfer functions.

To experiment with Layer 2 of the Grossberg network, use the Neural Network Design Demonstration Grossberg Layer 2 (`nnd15gl2`).

Learning Law

Long-Term Memory

The third component of the Grossberg network is the learning law for the adaptive weights $\mathbf{W}^2$. Grossberg calls these adaptive weights the *long-term memory* (LTM). This is because the rows of $\mathbf{W}^2$ will represent patterns that have been stored and that the network will be able to recognize. As in the Kohonen and Hamming networks, the stored pattern that is closest to an input pattern will produce the largest output in Layer 2. In the next subsection we will look more closely at the relationship between the Grossberg network and the Kohonen network.

One learning law for $\mathbf{W}^2$ is given by

$$\frac{dw^2_{i,j}(t)}{dt} = \alpha\{-w^2_{i,j}(t) + n^2_i(t)n^1_j(t)\} . \tag{18.24}$$

The first term in the bracket on the right-hand side of Eq. (18.24) is a passive decay term, which we have seen in the Layer 1 and Layer 2 equations, while the second term implements a Hebbian-type learning. Together, these terms implement the Hebb rule with decay, which was discussed in Chapter 13.

Recall from Chapter 13 that it is often useful to turn off learning (and forgetting) when $n^2_i(t)$ is not active. This can be accomplished by the following learning law:

$$\frac{dw^2_{i,j}(t)}{dt} = \alpha n^2_i(t)\{-w^2_{i,j}(t) + n^1_j(t)\} , \tag{18.25}$$

or, in vector form,

$$\frac{d[{}_i\mathbf{w}^2(t)]}{dt} = \alpha n^2_i(t)\{-[{}_i\mathbf{w}^2(t)] + \mathbf{n}^1(t)\} , \tag{18.26}$$

where ${}_i\mathbf{w}^2(t)$ is a vector composed of the elements of the ith row of $\mathbf{W}^2$ (see Eq. (4.4)).

The terms on the right-hand side of Eq. (18.25) are multiplied (gated) by $n^2_i(t)$, which allows learning (and forgetting) to occur only when $n^2_i(t)$ is not zero. This is the continuous-time implementation of the instar learning rule, which we introduced in Chapter 13 (Eq. (15.32)). In the following subsection we will demonstrate the equivalence of Eq. (18.25) and Eq. (15.32).

To illustrate the performance of the Grossberg learning law, consider a network with two neurons in each layer. The weight update equations would be

$$\frac{dw^2_{1,1}(t)}{dt} = n^2_1(t)\{-w^2_{1,1}(t) + n^1_1(t)\}, \tag{18.27}$$

$$\frac{dw^2_{1,2}(t)}{dt} = n^2_1(t)\{-w^2_{1,2}(t) + n^1_2(t)\}, \tag{18.28}$$

$$\frac{dw^2_{2,1}(t)}{dt} = n^2_2(t)\{-w^2_{2,1}(t) + n^1_1(t)\}, \tag{18.29}$$

$$\frac{dw^2_{2,2}(t)}{dt} = n^2_2(t)\{-w^2_{2,2}(t) + n^1_2(t)\}, \tag{18.30}$$

where the learning rate coefficient α has been set to 1. To simplify our example, we will assume that two different input patterns are alternately presented to the network for periods of 0.2 seconds at a time. We will also assume that Layer 1 and Layer 2 converge very quickly, in comparison with the convergence of the weights, so that the neuron outputs are effectively constant over the 0.2 seconds. The Layer 1 and Layer 2 outputs for the two different input patterns will be

$$\text{for pattern 1: } \mathbf{n}^1 = \begin{bmatrix} 0.9 \\ 0.45 \end{bmatrix}, \ \mathbf{n}^2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \tag{18.31}$$

$$\text{for pattern 2: } \mathbf{n}^1 = \begin{bmatrix} 0.45 \\ 0.9 \end{bmatrix}, \ \mathbf{n}^2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \tag{18.32}$$

Pattern 1 is coded by the first neuron in Layer 2, and pattern 2 is coded by the second neuron in Layer 2.

Figure 18.20 illustrates the response of the adaptive weights, beginning with all weights set to zero. Note that the first row of the weight matrix ($w^2_{1,1}(t)$ and $w^2_{1,2}(t)$) is only adjusted during those periods when $n^2_1(t)$ is nonzero, and that it converges to the corresponding $\mathbf{n}^1$ pattern ($n^1_1(t) = 0.9$ and $n^1_2(t) = 0.45$). (The elements in the first row of the weight matrix are indicated by the blue lines in Figure 18.20.) Also, the second row of the weight matrix ($w^2_{2,1}(t)$ and $w^2_{2,2}(t)$) is only adjusted during those periods when $n^2_2(t)$ is nonzero, and it converges to the corresponding $\mathbf{n}^1$ pattern ($n^1_1(t) = 0.45$ and $n^1_2(t) = 0.9$). (The elements in the second row of the weight matrix are indicated by the black lines in Figure 18.20.)

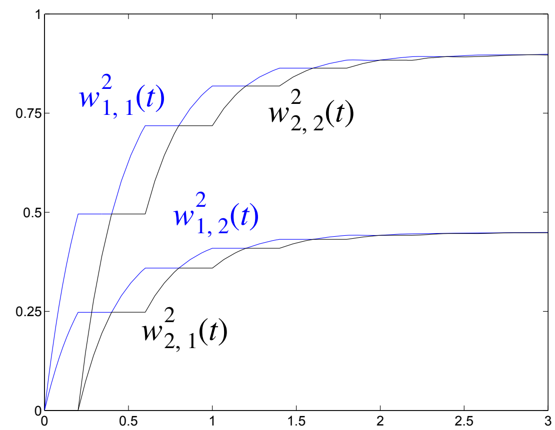


Figure 18.20 Response of the Adaptive Weights

To experiment with the adaptive weights, use the Neural Network Design Demonstration Adaptive Weights (`nnd15aw`).

Relation to Kohonen Law

In the previous section we indicated that the Grossberg learning law was a continuous-time version of the instar learning law, which we discussed in Chapter 13. Now we want to demonstrate this fact. We will also show that the Grossberg network is, in its simplest form, a continuous-time version of the Kohonen competitive network of Chapter 14.

To begin, let's repeat the Grossberg learning law of Eq. (18.25):

$$\frac{d[{}_i\mathbf{w}^2(t)]}{dt} = \alpha n_i^2(t)\{-[{}_i\mathbf{w}^2(t)] + \mathbf{n}^1(t)\} . \tag{18.33}$$

If we approximate the derivative by

$$\frac{d[{}_i\mathbf{w}^2(t)]}{dt} \approx \frac{{}_i\mathbf{w}^2(t+\Delta t) - {}_i\mathbf{w}^2(t)}{\Delta t} , \tag{18.34}$$

then we can rewrite Eq. (18.33) as

$${}_i\mathbf{w}^2(t+\Delta t) = {}_i\mathbf{w}^2(t) + \alpha(\Delta t) n_i^2(t)\{-{}_i\mathbf{w}^2(t) + \mathbf{n}^1(t)\} . \tag{18.35}$$

(Compare this equation with the instar rule that was presented in Chapter 13 in Eq. (15.33).) If we rearrange terms, this can be reduced to

$${}_i\mathbf{w}^2(t+\Delta t) = \{1 - \alpha(\Delta t) n_i^2(t)\}\, {}_i\mathbf{w}^2(t) + \alpha(\Delta t) n_i^2(t)\{\mathbf{n}^1(t)\} . \tag{18.36}$$

To simplify the analysis further, assume that a faster-than-linear transfer function is used in Layer 2, so that only one neuron in that layer will have a nonzero output; call it neuron $i*$. Then only row $i*$ of the weight matrix will be updated:

$$
{}_{i*}\mathbf{w}^2(t + \Delta t) = \{1 - \alpha'\}\,{}_{i*}\mathbf{w}^2(t) + \{\alpha'\}\mathbf{n}^1(t)\,, \tag{18.37}
$$

where $\alpha' = \alpha(\Delta t)n^2_{i*}(t)$.

This is almost identical to the Kohonen rule for the competitive network that we introduced in Chapter 14 in Eq. (16.13). The weight vector for the winning neuron (with nonzero output) will be moved toward $\mathbf{n}^1$, which is a normalized version of the current input pattern.

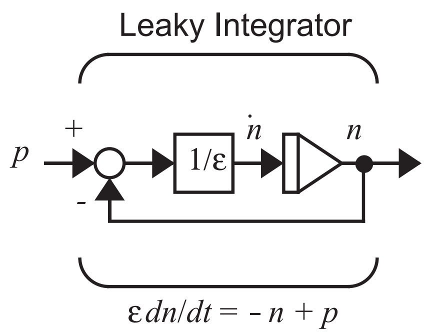
There are three major differences between the Grossberg network that we have presented in this chapter and the basic Kohonen competitive network. First, the Grossberg network is a continuous-time network (satisfies a set of nonlinear differential equations). Second, Layer 1 of the Grossberg network automatically normalizes the input vectors. Third, Layer 2 of the Grossberg network can perform a "soft" competition, rather than the winner-take-all competition of the Kohonen network. This soft competition allows more than one neuron in Layer 2 to learn. This causes the Grossberg network to operate as a feature map.

Summary of Results

Basic Nonlinear Model

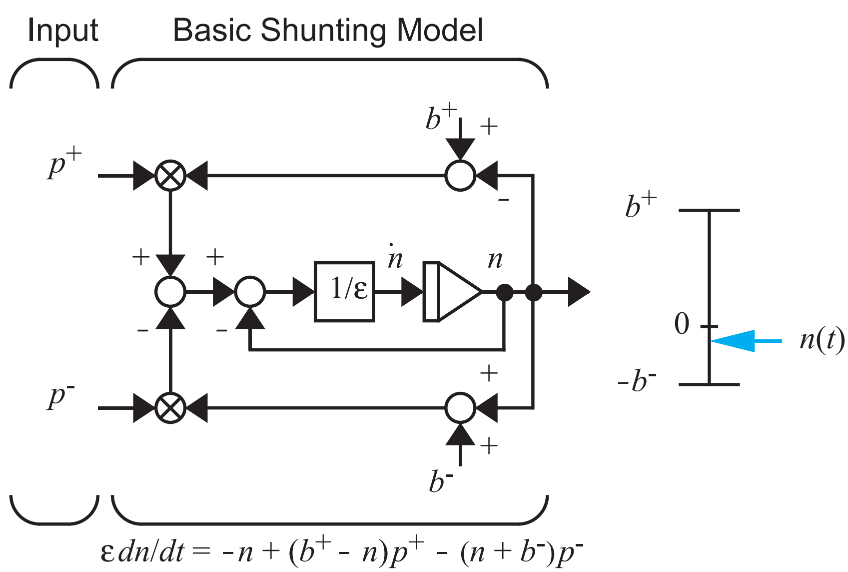
Leaky Integrator

$$\varepsilon\frac{dn(t)}{dt} = -n(t) + p(t)$$

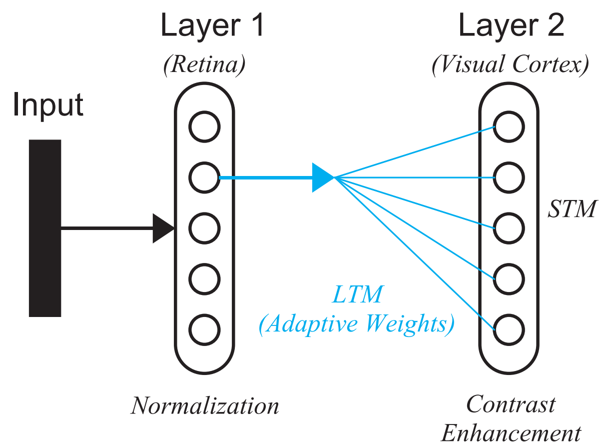


Shunting Model

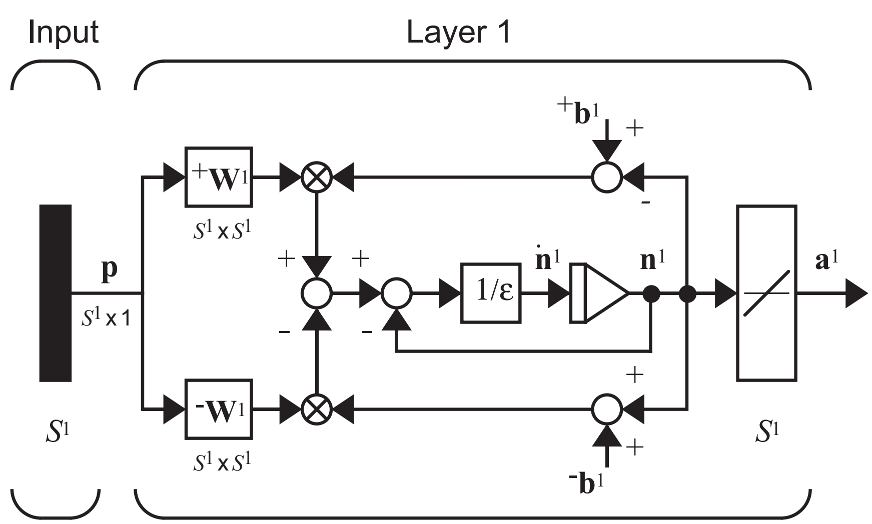
$$\varepsilon\frac{dn(t)}{dt} = -n(t) + (b^{+} - n(t))p^{+} - (n(t) + b^{-})p^{-}$$



Two-Layer Competitive Network



Layer 1



$$\varepsilon d\mathbf{n}^1/dt = -\mathbf{n}^1 + ({}^{+}\mathbf{b}^1 - \mathbf{n}^1)[{}^{+}\mathbf{W}^1]\mathbf{p} - (\mathbf{n}^1 + {}^{-}\mathbf{b}^1)[{}^{-}\mathbf{W}^1]\mathbf{p}$$

$$\varepsilon\frac{d\mathbf{n}^1(t)}{dt} = -\mathbf{n}^1(t) + ({}^{+}\mathbf{b}^1 - \mathbf{n}^1(t))[{}^{+}\mathbf{W}^1]\mathbf{p} - (\mathbf{n}^1(t) + {}^{-}\mathbf{b}^1)[{}^{-}\mathbf{W}^1]\mathbf{p}$$

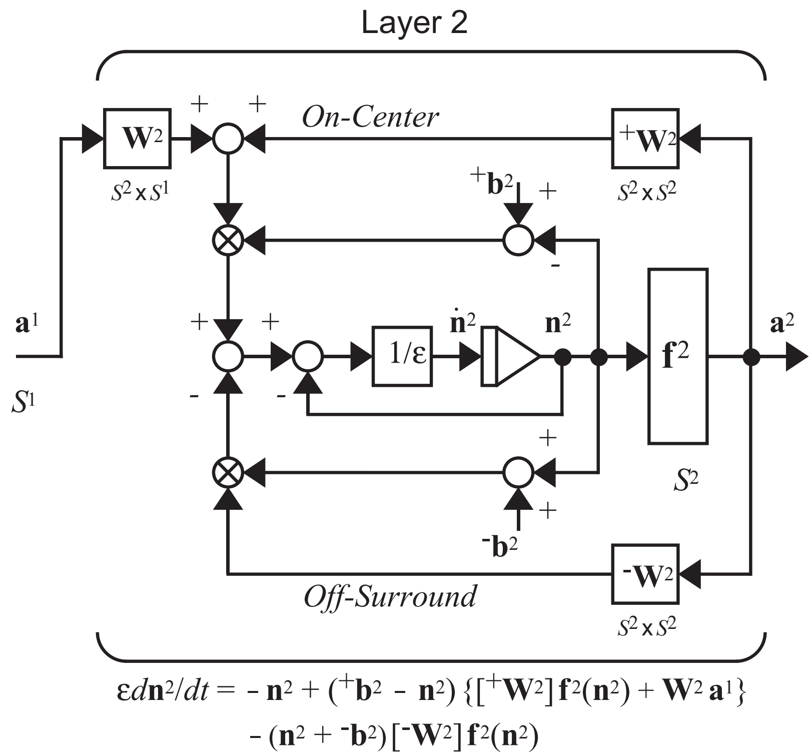
$${}^{+}\mathbf{W}^1 = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \qquad {}^{-}\mathbf{W}^1 = \begin{bmatrix} 0 & 1 & \cdots & 1 \\ 1 & 0 & \cdots & 1 \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \cdots & 0 \end{bmatrix}$$

On-Center *Off-Surround*

Steady State Neuron Activity

$$n_i^1 = \left(\frac{{}^{+}b^1 P}{1+P}\right)\bar{p}_i \text{, where } \bar{p}_i = \frac{p_i}{P} \text{ and } P = \sum_{j=1}^{S^1} p_j$$

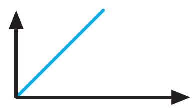
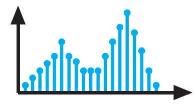
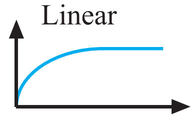
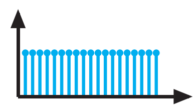
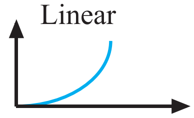
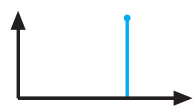
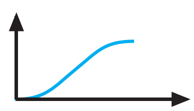
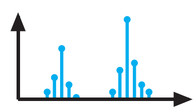
Layer 2



$$\varepsilon\frac{d\mathbf{n}^2(t)}{dt} = -\mathbf{n}^2(t) + ({}^{+}\mathbf{b}^2 - \mathbf{n}^2(t))\{[{}^{+}\mathbf{W}^2]\mathbf{f}^2(\mathbf{n}^2(t)) + \mathbf{W}^2\mathbf{a}^1\}$$

$$-(\mathbf{n}^2(t) + {}^{-}\mathbf{b}^2)[{}^{-}\mathbf{W}^2]\mathbf{f}^2(\mathbf{n}^2(t))$$

Choice of Transfer Function

$f^2(n)$	Stored Pattern $\mathbf{n}^2(\infty)$	Comments
Linear 		Perfect storage of any pattern, but amplifies noise.
Slower than Linear 		Amplifies noise, reduces contrast.
Faster than Linear 		Winner-take-all, suppresses noise, quantizes total activity.
Sigmoid 		Supresses noise, contrast enhances, not quantized.

Learning Law

$$\frac{d[{}_i\mathbf{w}^2(t)]}{dt} = \alpha n_i^2(t)\{-[{}_i\mathbf{w}^2(t)] + \mathbf{n}^1(t)\}$$

(Continuous-Time Instar Learning)

Solved Problems

P18.1 Demonstrate the effect of the coefficient ε on the performance of the leaky integrator, which is shown in Figure P18.1, with the input $p = 1$.

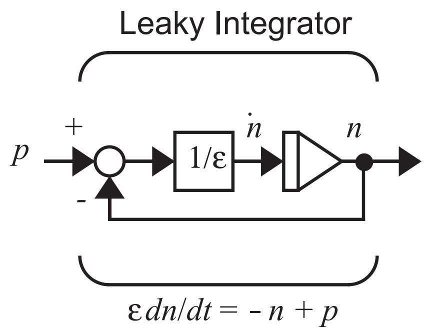


Figure P18.1 Leaky Integrator

The equation of operation for the leaky integrator is

$$\varepsilon \frac{dn(t)}{dt} = -n(t) + p(t) .$$

The solution to this differential equation, for an arbitrary input $p(t)$, is

$$n(t) = e^{-t/\varepsilon} n(0) + \frac{1}{\varepsilon} \int_0^t e^{-(t-\tau)/\varepsilon} p(t-\tau) d\tau .$$

If $p(t) = 1$, the solution will be

$$n(t) = e^{-t/\varepsilon} n(0) + \frac{1}{\varepsilon} \int_0^t e^{-(t-\tau)/\varepsilon} d\tau .$$

We want to show how this response changes as a function of ε. The response will be

$$n(t) = e^{-t/\varepsilon} n(0) + (1 - e^{-t/\varepsilon}) = e^{-t/\varepsilon}(n(0) - 1) + 1 .$$

This response begins at $n(0)$, and then grows exponentially (or decays exponentially, depending on whether or not $n(0)$ is greater than or less than 1), approaching the steady state response of $n(\infty) = 1$. As ε is decreased, the response becomes faster (since $e^{-t/\varepsilon}$ decays more quickly), while the steady state value remains constant. Figure P18.2 illustrates the responses for $\varepsilon = 1, 0.5, 0.25, 0.125$, with $n(0) = 0$. Notice that the steady state value remains 1 for each case. Only the speed of response changes.

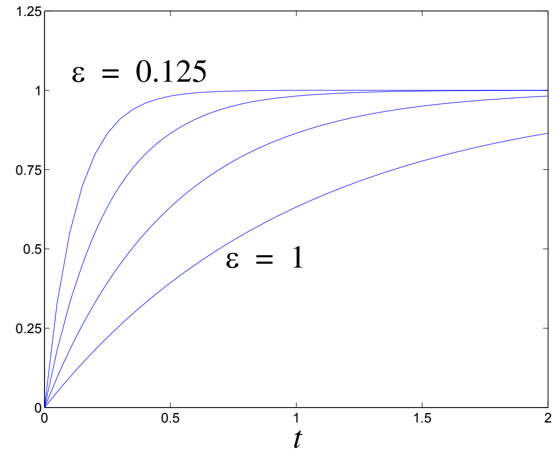


Figure P18.2 Effect of ε on Leaky Integrator Response

P18.2 Again using the leaky integrator of Figure P18.1, set $\varepsilon = 1$.

i. Find a difference equation approximation to the leaky integrator differential equation by approximating the derivative using

$$\frac{dn(t)}{dt} \approx \frac{n(t + \Delta t) - n(t)}{\Delta t} .$$

ii. Using $\Delta t = 0.1$, compare the response of this difference equation with the response of the differential equation for $p(t) = 1$ and $n(0) = 0$. Compare the two over the range $0 < t < 1$.

iii. Using the difference equation model for the leaky integrator, show that the response is a weighted average of previous inputs.

i. If we make the approximation to the derivative, we find

$$\frac{n(t + \Delta t) - n(t)}{\Delta t} = -n(t) + p(t)$$

or

$$n(t + \Delta t) = n(t) + \Delta t\{-n(t) + p(t)\} = (1 - \Delta t)n(t) + (\Delta t)p(t) .$$

ii. If we let $\Delta t = 0.1$ we obtain the difference equation

$$n(t + 0.1) = 0.9n(t) + 0.1p(t) .$$

If we let $p(t) = 1$ and $n(0) = 0$, then we can solve for $n(t)$:

$$n(0.1) = 0.9n(0) + 0.1p(0) = 0.1$$

$$n(0.2) = 0.9n(0.1) + 0.1p(0.1) = 0.9(0.1) + 0.1(1) = 0.19,$$

$$n(0.3) = 0.9n(0.2) + 0.1p(0.2) = 0.9(0.19) + 0.1(1) = 0.271,$$

$$n(0.4) = 0.9n(0.3) + 0.1p(0.3) = 0.9(0.271) + 0.1(1) = 0.3439,$$

$$n(0.5) = 0.9n(0.4) + 0.1p(0.4) = 0.9(0.3439) + 0.1(1) = 0.4095,$$

$$n(0.6) = 0.4686, \; n(0.7) = 0.5217, \; n(0.8) = 0.5695,$$

$$n(0.9) = 0.6126, \; n(1.0) = 0.6513.$$

From Problem P18.1, the solution to the differential equation is

$$n(t) = e^{-t/\varepsilon}n(0) + (1 - e^{-t/\varepsilon}) = (1 - e^{-t}).$$

Figure P18.3 illustrates the relationship between the difference equation solution and the differential equation solution. The black line represents the differential equation solution, and the blue circles represent the difference equation solution. The two solutions are very close, and can be made arbitrarily close by decreasing the interval Δt.

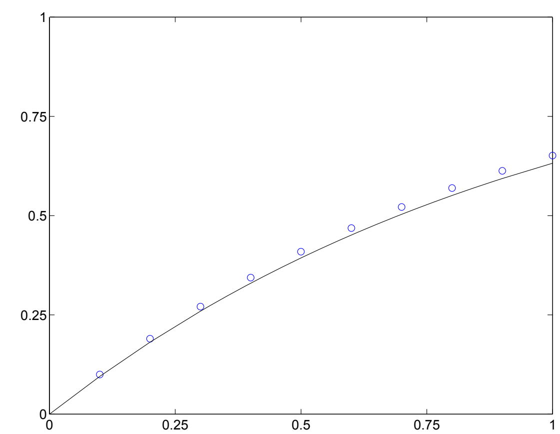


Figure P18.3 Comparison of Difference and Differential Equations

iii. Consider again the difference equation model of the leaky integrator, which we developed in part (ii):

$$n(t + 0.1) = 0.9n(t) + 0.1p(t).$$

If we start from a zero initial condition we find

$$n(0.1) = 0.9n(0) + 0.1p(0) = 0.1p(0),$$

$$n(0.2) = 0.9n(0.1) + 0.1p(0.1) = 0.9\{0.1p(0)\} + 0.1p(0.1) = 0.09p(0) + 0.1p(0.1)$$

$$n(0.3) = 0.9n(0.2) + 0.1p(0.2) = 0.081p(0) + 0.09p(0.1) + 0.1p(0.2)$$

$$\vdots$$

$$n(k0.1) = 0.1\{(0.9)^{k-1}p(0) + (0.9)^{k-2}p(0.1) + \cdots + p((k-1)0.1)\}.$$

Therefore the response of the leaky integrator is a weighted average of previous inputs, $p(0), p(0.1), \ldots, p((k-1)0.1)$. Note that the recent inputs contribute more to the response than the early inputs.

P18.3 Find the response of the shunting network shown in Figure P18.4 for $\varepsilon = 1$, $b^+ = 1$, $b^- = 1$, $p^+ = 0$, $p^- = 10$ and $n(0) = 0.5$.

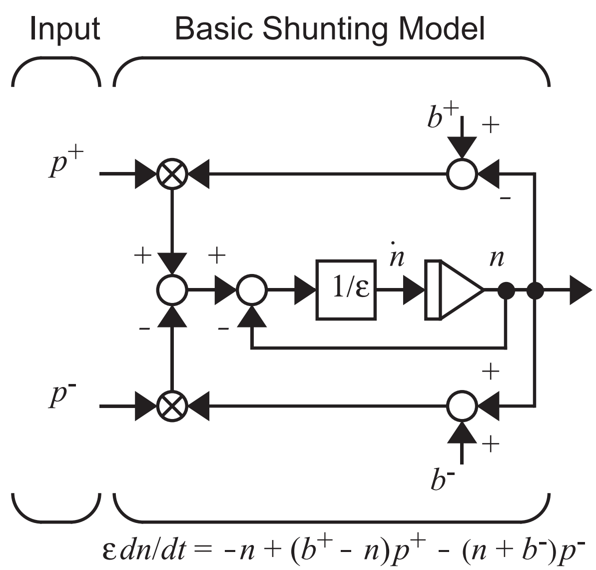


Figure P18.4 Shunting Network

The equation of operation of the shunting network is

$$\varepsilon\frac{dn(t)}{dt} = -n(t) + (b^+ - n(t))p^+ - (n(t) + b^-)p^-.$$

For the given parameter values this becomes

$$\frac{dn(t)}{dt} = -n(t) - (n(t) + 1)10 = -11n(t) - 10.$$

The solution to this equation is

$$n(t) = e^{-11t}n(0) + \int_0^t e^{-11(t-\tau)}(-10)d\tau,$$

or

$$n(t) = e^{-11t}0.5 + \left(-\frac{10}{11}\right)(1 - e^{-11t}).$$

The response is plotted in Figure P18.5.

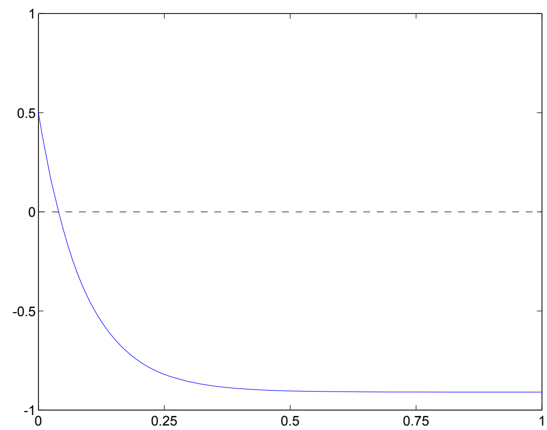


Figure P18.5 Shunting Network Response

There are two things to note about this response. First, as with all shunting networks, the response will never drop below $-{}^{-}b$, which in this case is -1. As the inhibitory input ${}^{-}p$ is increased, the steady state response will decrease, but it can never be less than $-{}^{-}b$. The second characteristic of the response is that the speed of the response will increase as the input is increased. For instance, if the input were changed from ${}^{-}p = 10$ to ${}^{-}p = 100$, the response would be

$$n(t) = e^{-101t}0.5 + \left(-\frac{100}{101}\right)(1 - e^{-101t}).$$

Since e^{-101t} decays more rapidly than e^{-11t}, the response will be faster.

P18.4 Find the response of Layer 1 of the Grossberg network for the case of two neurons, with ${}^{+}b^1 = 1$, ${}^{-}b^1 = 0$, $\varepsilon = 1$ and input vector $\mathbf{p} = \begin{bmatrix} c & 2c \end{bmatrix}^T$. Assume that the initial conditions are set to zero. Demonstrate the effect of c on the response.

The Layer 1 differential equations for this case are

$$\frac{dn_1^1(t)}{dt} = -n_1^1(t) + (1 - n_1^1(t))(c) - n_1^1(t)(2c) = -(1 + 3c)n_1^1(t) + c,$$

$$\frac{dn_2^1(t)}{dt} = -n_2^1(t) + (1 - n_2^1(t))(2c) - n_2^1(t)(c) = -(1+3c)n_2^1(t) + 2c .$$

The solutions to these equations would be

$$n_1^1(t) = e^{-(1+3c)t}n_1^1(0) + \int_0^t e^{-(1+3c)(t-\tau)}(c)d\tau ,$$

$$n_2^1(t) = e^{-(1+3c)t}n_2^1(0) + \int_0^t e^{-(1+3c)(t-\tau)}(2c)d\tau .$$

If the initial conditions are set to zero, these equations reduce to

$$n_1^1(t) = \left(\frac{c}{1+3c}\right)(1 - e^{-(1+3c)t}) ,$$

$$n_2^1(t) = \left(\frac{2c}{1+3c}\right)(1 - e^{-(1+3c)t}) .$$

Note that the outputs of Layer 1 retain the same relative intensities as the inputs; the output of neuron 2 is always twice the output of neuron 1. This behavior is consistent with Eq. (18.13). In addition, the total output intensity $(n_1^1(t) + n_2^1(t))$ is never larger than ${}^{+}b^1 = 1$, as predicted in Eq. (18.14).

As c is increased, it has two effects on the response. First, the steady state values increase slightly. Second, the response becomes faster, since $e^{-(1+3c)t}$ decays more rapidly as c increases.

P18.5 Consider Layer 2 of the Grossberg network. Assume that the input to Layer 2 is applied for some length of time and then removed (set to zero).

i. Find a differential equation that describes the variation in the total output of Layer 2,

$$N^2(t) = \sum_{k=1}^{S^2} n_k^2(t) ,$$

after the input to Layer 2 has been removed.

ii. Find a differential equation that describes the variation in the relative outputs of Layer 2,

$$\bar{n}_i^2(t) = \frac{n_i^2(t)}{N^2(t)} ,$$

after the input to Layer 2 has been removed.

i. The operation of Layer 2 is described by Eq. (18.17):

$$\varepsilon \frac{d\mathbf{n}^2(t)}{dt} = -\mathbf{n}^2(t) + ({}^{+}\mathbf{b}^2 - \mathbf{n}^2(t))\{[{}^{+}\mathbf{W}^2]\mathbf{f}^2(\mathbf{n}^2(t)) + \mathbf{W}^2\mathbf{a}^1\}$$
$$- (\mathbf{n}^2(t) + {}^{-}\mathbf{b}^2)[{}^{-}\mathbf{W}^2]\mathbf{f}^2(\mathbf{n}^2(t)) .$$

If the input is removed, then $\mathbf{W}^2\mathbf{a}^1$ is zero. For simplicity, we will set the inhibitory bias ${}^{-}\mathbf{b}^2$ to zero, and we will set all elements of the excitatory bias ${}^{+}\mathbf{b}^2$ to ${}^{+}b^2$. The response of neuron i is then given by

$$\varepsilon \frac{dn_i^2(t)}{dt} = -n_i^2(t) + ({}^{+}b^2 - n_i^2(t))\{f^2(n_i^2(t))\} - n_i^2(t)\left\{\sum_{k \neq i} f^2(n_k^2(t))\right\}.$$

This can be rearranged to produce

$$\varepsilon \frac{dn_i^2(t)}{dt} = -n_i^2(t) + {}^{+}b^2\{f^2(n_i^2(t))\} - n_i^2(t)\left\{\sum_{k=1}^{S^2} f^2(n_k^2(t))\right\}.$$

If we then make the definition

$$F^2(t) = \sum_{k=1}^{S^2} f^2(n_k^2(t)) ,$$

we can simplify the equation to

$$\varepsilon \frac{dn_i^2(t)}{dt} = -(1 + F^2(t))n_i^2(t) + {}^{+}b^2\{f^2(n_i^2(t))\} .$$

To get the total activity, sum this equation over i to produce

$$\varepsilon \frac{dN^2(t)}{dt} = -(1 + F^2(t))N^2(t) + {}^{+}b^2\{F^2(t)\} .$$

This equation describes the variation in the total activity in Layer 2 over time.

ii. The derivative of the relative activity is

$$\frac{d}{dt}[\bar{n}_i^2(t)] = \frac{d}{dt}\left[\frac{n_i^2(t)}{N^2(t)}\right] = \frac{1}{N^2(t)}\frac{d}{dt}[n_i^2(t)] - \left[\frac{n_i^2(t)}{(N^2(t))^2}\right]\frac{d}{dt}[N^2(t)] .$$

If we then substitute our previous equations for these derivatives, we find

$$\varepsilon\frac{d}{dt}[\bar{n}_i^2(t)] = \frac{1}{N^2(t)}\Bigg[\{-(1+F^2(t))n_i^2(t) + {}^+b^2\{f^2(n_i^2(t))\}\}$$
$$-\frac{n_i^2(t)}{N^2(t)}\{-(1+F^2(t))N^2(t) + {}^+b^2\{F^2(t)\}\}\Bigg] .$$

Two terms on the right-hand side will cancel to produce

$$\varepsilon\frac{d}{dt}[\bar{n}_i^2(t)] = \frac{1}{N^2(t)}\left[\{{}^+b^2\{f^2(n_i^2(t))\}\} - \frac{n_i^2(t)}{N^2(t)}\{{}^+b^2\{F^2(t)\}\}\right],$$

or

$$\varepsilon\frac{d}{dt}[\bar{n}_i^2(t)] = \frac{{}^+b^2F^2(t)}{N^2(t)}\left[\frac{f^2(n_i^2(t))}{F^2(t)} - \frac{n_i^2(t)}{N^2(t)}\right].$$

We can put this in a more useful form if we expand the terms in the brackets:

$$\left[\frac{f^2(n_i^2(t))}{F^2(t)} - \frac{n_i^2(t)}{N^2(t)}\right] = \frac{1}{F^2(t)N^2(t)}[f^2(n_i^2(t))N^2(t) - n_i^2(t)F^2(t)]$$

$$= \frac{1}{F^2(t)N^2(t)}\left[g^2(n_i^2(t))n_i^2(t)\sum_{k=1}^{S^2} n_k^2(t) - n_i^2(t)\sum_{k=1}^{S^2} g^2(n_k^2(t))n_k^2(t)\right]$$

$$= \frac{n_i^2(t)}{F^2(t)N^2(t)}\left[\sum_{k=1}^{S^2} n_k^2(t)[g^2(n_i^2(t)) - g^2(n_k^2(t))]\right],$$

where

$$g^2(n_i^2(t)) = \frac{f^2(n_i^2(t))}{n_i^2(t)} .$$

Combining this expression with our previous equation, we obtain

$$\varepsilon \frac{d}{dt}[\bar{n}_i^2(t)] = {}^{+}b^2 \bar{n}_i^2(t) \left[\sum_{k=1}^{S^2} \bar{n}_k^2(t) [g^2(n_i^2(t)) - g^2(n_k^2(t))] \right].$$

This form of the differential equation describing the evolution of the relative outputs is very useful in demonstrating the characteristics of Layer 2, as we will see in the next solved problem.

P18.6 Suppose that the transfer function in Layer 2 of the Grossberg network is linear.

i. Show that the relative outputs of Layer 2 will not change after the input has been removed.

ii. Under what conditions will the total output of Layer 2 decay to zero after the input has been removed?

i. From Problem P18.5 we know that the relative outputs of Layer 2, after the input has been removed, evolve according to

$$\varepsilon \frac{d}{dt}[\bar{n}_i^2(t)] = {}^{+}b^2 \bar{n}_i^2(t) \left[\sum_{k=1}^{S^2} \bar{n}_k^2(t) [g^2(n_i^2(t)) - g^2(n_k^2(t))] \right].$$

If the transfer function for Layer 2, $f^2(n)$, is linear, then

$$f^2(n) = c\, n .$$

Therefore

$$g^2(n) = \frac{f^2(n)}{n} = \frac{c\, n}{n} = c .$$

If we substitute this expression into our differential equation, we find

$$\varepsilon \frac{d}{dt}[\bar{n}_i^2(t)] = {}^{+}b^2 \bar{n}_i^2(t) \left[\sum_{k=1}^{S^2} \bar{n}_k^2(t) [c - c] \right] = 0 .$$

Therefore the relative outputs do not change.

ii. From Problem P18.5, the total output of Layer 2, after the input has been removed, evolves according to

$$\varepsilon \frac{dN^2(t)}{dt} = -(1 + F^2(t)) N^2(t) + {}^{+}b^2 \{F^2(t)\} .$$

If $f^2(n)$ is linear, then

$$F^2(t) = \sum_{k=1}^{S^2} f^2(n_k^2(t)) = \sum_{k=1}^{S^2} c\, n_k^2(t) = c \sum_{k=1}^{S^2} n_k^2(t) = c\, N^2(t) .$$

Therefore the differential equation can be written

$$\varepsilon \frac{dN^2(t)}{dt} = -(1 + c\, N^2(t))N^2(t) + {}^+b^2\{c\, N^2(t)\} = -\{1 - {}^+b^2 c + c\, N^2(t)\}N^2(t) .$$

To find the equilibrium solutions of this equation, we set the derivative to zero:

$$0 = -\{1 - {}^+b^2 c + c\, N^2(t)\}N^2(t) .$$

Therefore there are two equilibrium solutions:

$$N^2(t) = 0 \text{ or } N^2(t) = \frac{{}^+b^2 c - 1}{c} .$$

We want to know the conditions under which the total output will converge to each of these possible solutions. Consider two cases:

1. $1 \geq {}^+b^2 c$

   For this case, the derivative of the total output,

   $$\varepsilon \frac{dN^2(t)}{dt} = -\{1 - {}^+b^2 c + c\, N^2(t)\}N^2(t) ,$$

   will always be negative for positive $N^2(t)$. (Recall that the outputs of Layer 2 are never negative.) Therefore, the total output will decay to zero.

   $$\lim_{t \to \infty} N^2(t) = 0$$

2. $1 < {}^+b^2 c$

   (a) If $N^2(0) > ({}^+b^2 c - 1)/c$, then the derivative of the total output will be negative until $N^2(t) = ({}^+b^2 c - 1)/c$, when the derivative will be zero. Therefore,

   $$\lim_{t \to \infty} N^2(t) = \frac{({}^+b^2 c - 1)}{c} .$$

(b) If $N^2(0) < ({}^{+}b^2 c - 1)/c$, then the derivative of the total output will be positive until $N^2(t) = ({}^{+}b^2 c - 1)/c$, when the derivative will be zero. Therefore,

$$\lim_{t \to \infty} N^2(t) = \frac{({}^{+}b^2 c - 1)}{c}.$$

Therefore, if the transfer function of Layer 2 is linear, the total output will decay to zero if $1 \geq {}^{+}b^2 c$. If $1 < {}^{+}b^2 c$, then the total output will converge to $({}^{+}b^2 c - 1)/c$. In any case, the relative outputs will remain constant.

As an example of these results, consider the following Layer 2 equations:

$$\frac{dn_1^2(t)}{dt} = -n_1^2(t) + (1.5 - n_1^2(t))\{n_1^2(t)\} - n_1^2(t)\{n_2^2(t)\},$$

$$\frac{dn_2^2(t)}{dt} = -n_2^2(t) + (1.5 - n_2^2(t))\{n_2^2(t)\} - n_2^2(t)\{n_1^2(t)\}.$$

For this case, $\varepsilon = 1$, ${}^{+}b^2 = 1.5$ and $c = 1$, therefore $1 < {}^{+}b^2 c$. The total output will converge to

$$\lim_{t \to \infty} N^2(t) = \frac{({}^{+}b^2 c - 1)}{c} = \frac{(1.5 - 1)}{1} = 0.5.$$

In Figure P18.6 we can see the response of Layer 2 for two different sets of initial conditions:

$$\mathbf{n}^2(0) = \begin{bmatrix} 0.75 \\ 0.5 \end{bmatrix} \text{ and } \mathbf{n}^2(0) = \begin{bmatrix} 0.15 \\ 0.1 \end{bmatrix}.$$

As expected, the total output converges to 0.5 for both initial conditions. In addition, since the relative values of the initial conditions are the same for the two cases, the outputs converge to the same values in both cases.

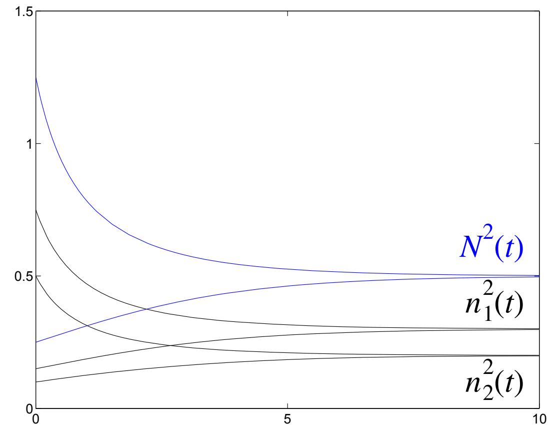


Figure P18.6 Response of Layer 2 for Linear $f^2(n)$

P18.7 Show that the continuous-time Hebb rule with decay, given by Eq. (18.24), is equivalent to the discrete-time version given by Eq. (15.18).

The continuous-time Hebb rule with decay is

$$\frac{dw_{i,j}^2(t)}{dt} = \alpha\{-w_{i,j}^2(t) + n_i^2(t)n_j^1(t)\}.$$

If we approximate the derivative by

$$\frac{dw_{i,j}^2(t)}{dt} \approx \frac{w_{i,j}^2(t+\Delta t) - w_{i,j}^2(t)}{\Delta t},$$

the Hebb rule becomes

$$w_{i,j}^2(t+\Delta t) = w_{i,j}^2(t) + \alpha\Delta t\{-w_{i,j}^2(t) + n_i^2(t)n_j^1(t)\}.$$

This can be rearranged to obtain

$$w_{i,j}^2(t+\Delta t) = [1-\alpha\Delta t]w_{i,j}^2(t) + \alpha\Delta t\{n_i^2(t)n_j^1(t)\}.$$

In vector form this would be

$$\mathbf{W}^2(t+\Delta t) = [1-\alpha\Delta t]\mathbf{W}^2(t) + \alpha\Delta t\{\mathbf{n}^2(t)(\mathbf{n}^1(t))^T\}.$$

If we compare this with Eq. (15.18),

$$\mathbf{W}(q) = (1-\gamma)\mathbf{W}(q-1) + \alpha\mathbf{a}(q)\mathbf{p}^T(q),$$

we can see that they have the identical form.

Epilogue

The Grossberg network presented in this chapter was inspired by the visual system of higher vertebrates. To motivate the network, we presented a brief description of the primary visual pathway. We also discussed some visual illusions, which help us to understand the mechanisms underlying the visual system.

The Grossberg network is a two-layer, continuous-time competitive network, which is very similar in structure and operation to the Kohonen competitive network presented in Chapter 14. The first layer of the Grossberg network normalizes the input pattern. It demonstrates how the visual system can use on-center/off-surround connection patterns and a shunting model to implement an automatic gain control, which normalizes total activity.

The second layer of the Grossberg network performs a competition, which contrast enhances the output pattern and stores it in short-term memory. It uses nonlinear feedback and the on-center/off-surround connection pattern to produce the competition and the storage. The choice of the transfer function and the feedback connection pattern determines the degree of competition (e.g., winner-take-all, mild contrast enhancement, or no change in the pattern).

The adaptive weights in the Grossberg network use an instar learning rule, which stores prototype patterns in long-term memory. When a winner-take-all competition is performed in the second layer, this learning rule is equivalent to the Kohonen learning rule used in Chapter 14.

As with the Kohonen network, one key problem of the Grossberg network is the stability of learning; as more inputs are applied to the network, the weight matrix may never converge. This problem was discussed extensively in Chapter 14. In Chapter 16 we will present a class of networks that is designed to overcome this difficulty: the Adaptive Resonance Theory (ART) networks. The ART networks are direct descendents of the Grossberg network presented in this chapter.

Another problem with the Grossberg network, which we have not discussed in this chapter, is the stability of the differential equations that implement the network. In Layer 2, for example, we have a set of differential equations with nonlinear feedback. Can we make some general statement about the stability of such systems? Chapter 17 will present a comprehensive discussion of this problem.

Further Reading

[GrMi89] S. Grossberg, E. Mingolla and D. Todorovic, "A neural network architecture for preattentive vision," *IEEE Transactions on Biomedical Engineering*, vol. 36, no. 1, pp. 65–84, 1989.

The objective of this paper is to develop a neural network for general purpose preattentive vision. The network consists of two main subsystems: a boundary contour system and a feature contour system.

[Gros76] S. Grossberg, "Adaptive pattern classification and universal recoding: I. Parallel development and coding of neural feature detectors," *Biological Cybernetics*, vol. 23, pp. 121–134, 1976.

Grossberg describes a continuous-time competitive network, inspired by the developmental physiology of the visual cortex. The structure of this network forms the foundation for other important networks.

[Gros82] S. Grossberg, *Studies of Mind and Brain*, Boston: D. Reidel Publishing Co., 1982.

This book is a collection of Stephen Grossberg papers from the period 1968 through 1980. It covers many of the fundamental concepts that are used in later Grossberg networks, such as the adaptive resonance theory networks.

[Hube88] D.H. Hubel, *Eye, Brain, and Vision*, New York: Scientific American Library, 1988.

David Hubel has been at the center of research in this area for 30 years, and his book provides an excellent introduction to the human visual system. He explains the current view of the visual system in a way that is easily accessible to anyone with some scientific training.

[vanT75] H. F. J. M. van Tuijl, "A new visual illusion: Neonlike color spreading and complementary color induction between subjective contours," *Acta Psychologica*, vol. 39, pp. 441–445, 1975.

This paper describes the original discovery of the illusion in which crosses of certain colors, when placed inside Ehrenstein figures, appear to spread into solid shapes.

[vond73] C. von der Malsburg, "Self-organization of orientation sensitive cells in the striate cortex," *Kybernetic*, vol. 14, pp. 85–100, 1973.

Malsberg's is one of the first papers to present a self-organizing feature map neural network. The network is a model for the visual cortex of higher vertebrates. This paper influenced the work of Kohonen and Grossberg on feature maps.

Exercises

E18.1 Consider the leaky integrator shown in Figure E18.1.

i. Find the response $n(t)$ if $\varepsilon = 1$, $n(0) = 1$ and $p(t) = 0.5$.

ii. Find the response $n(t)$ if $\varepsilon = 1$, $n(0) = 1$ and $p(t) = 2$.

iii. Find the response $n(t)$ if $\varepsilon = 4$, $n(0) = 1$ and $p(t) = 2$.

iv. Check your answers to the previous parts by writing a MATLAB M-file to simulate the leaky integrator. Use the `ode45` routine. Plot the response for each case.

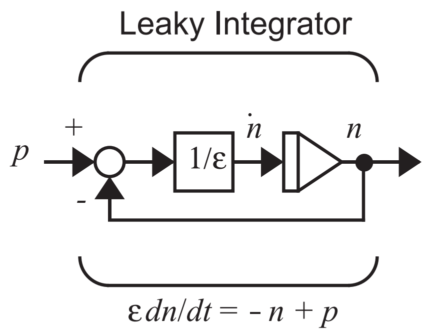


Figure E18.1 Leaky Integrator

E18.2 Consider the shunting network shown in Figure E18.2.

i. Find and sketch the response of the shunting network if $\varepsilon = 2$, $b^{+} = 3$, $b^{-} = 1$, $p^{+} = 0$, $p^{-} = 5$ and $n(0) = 1$.

ii. Find and sketch the response of the shunting network if $\varepsilon = 2$, $b^{+} = 3$, $b^{-} = 1$, $p^{+} = 0$, $p^{-} = 50$ and $n(0) = 1$.

iii. Find and sketch the response of the shunting network if $\varepsilon = 2$, $b^{+} = 3$, $b^{-} = 1$, $p^{+} = 50$, $p^{-} = 0$ and $n(0) = 1$.

iv. Find and sketch the response of the shunting network if $\varepsilon = 5$, $b^{+} = 2$, $b^{-} = 6$, $p^{+} = 5$, $p^{-} = 0$ and $n(0) = 0$.

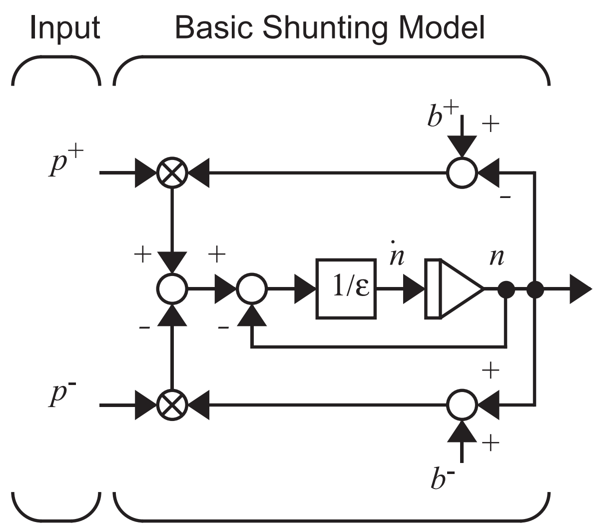
v. Find and sketch the response of the shunting network if $\varepsilon = 5$, $b^{+} = 2$, $b^{-} = 6$, $p^{+} = 0$, $p^{-} = 5$ and $n(0) = 0$.

vi. Find and sketch the response of the shunting network if $\varepsilon = 0.25$, $b^{+} = 4$, $b^{-} = 2$, $p^{+} = 2$, $p^{-} = 2$ and $n(0) = 0$.

vii. Find and sketch the response of the shunting network if $\varepsilon = 0.25$, $b^{+} = 4$, $b^{-} = 2$, $p^{+} = 2$, $p^{-} = 4$ and $n(0) = 0$.

viii. Check your answers to the previous parts by writing a MATLAB M-file to simulate the shunting network. Use the `ode45` routine. Plot the response for each case. Verify that your responses agree with the known characteristics of the shunting model.

ix. Explain the differences in operation of the leaky integrator and the shunting network.



$$\varepsilon\, dn/dt = -n + (b^+ - n)p^+ - (n + b^-)p^-$$

Figure E18.2 Shunting Network

E18.3 Suppose that Layer 1 of the Grossberg network has two neurons, with ${}^{+}b^1 = 0.5$, $\varepsilon = 0.5$ and input vector $\mathbf{p} = \begin{bmatrix} 2 & 1 \end{bmatrix}^T$. Assume that the initial conditions are set to zero.

i. Find the steady state response of Layer 1, using Eq. (18.13).

ii. Find the solution to the differential equation for Layer 1. Verify that the steady state response agrees with your answer to part (i).

iii. Check your answer by writing a MATLAB M-file to simulate Layer 1 of the Grossberg network. Use the `ode45` routine. Plot the response.

E18.4 Repeat Exercise E18.3 for input vector $\mathbf{p} = \begin{bmatrix} 20 & 10 \end{bmatrix}^T$.

E18.5 Consider the first layer of the Grossberg network. The parameters are set to be ${}^{+}b^1 = 2$, ${}^{-}b^1 = 0$, $\varepsilon = 2$. The input to the network is $\mathbf{p} = \begin{bmatrix} 4 & 1 \end{bmatrix}^T$. Find the first layer outputs and sketch them versus time.

E18.6 Find the differential equation that describes the variation in the total output of Layer 1,

$$N^1(t) = \sum_{i=1}^{S^1} n_i^1(t).$$

(Use the technique presented in Problem P18.5.)

E18.7 Assume that Layer 2 of the Grossberg network has two neurons, with $f^2(n) = 2n$, $\varepsilon = 1$, ${}^{+}b^2 = 1$ and ${}^{-}b^2 = 0$. The inputs have been applied for some length of time, then removed.

i. What will be the steady state total output, $\lim_{t \to \infty} N^2(t)$?

ii. Repeat part (i) if ${}^{+}b^2 = 0.25$.

iii. Check your answers to the previous parts by writing a MATLAB M-file to simulate Layer 2 of the Grossberg network. Use the `ode45` routine. Plot the responses for the following initial conditions:

$$\mathbf{n}^2(0) = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \text{ and } \mathbf{n}^2(0) = \begin{bmatrix} 0.2 \\ 0.1 \end{bmatrix}.$$

E18.8 Suppose that the transfer function for Layer 2 of the Grossberg network is $f^2(n) = c \times (n)^2$, and $\varepsilon = 1$, ${}^{+}b^2 = 1$.

i. Using the results of Problem P18.5, show that, after the inputs have been removed, all of the relative outputs of Layer 2 will decay to zero, except the one with the largest initial condition (winner-take-all competition).

ii. For what values of c will the total output $N^2(t)$ have a nonzero stable point (steady state value)?

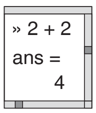
iii. If the condition of part (ii) is satisfied, what will be the steady state value of $N^2(t)$? Will this depend on the initial condition $N^2(0)$?

iv. Check your answers to the previous parts by writing a MATLAB M-file and simulating the total response of Layer 2 for $c = 4$ and $N^2(0) = 3$.

E18.9 Simulate the response of the adaptive weights for the Grossberg network. Assume that the coefficient ε is 1. Assume that two different input patterns are alternately presented to the network for periods of 0.2 seconds at a time. Also, assume that Layer 1 and Layer 2 converge very quickly, in comparison with the convergence of the weights, so that the neuron outputs are effectively constant over the 0.2 seconds. The Layer 2 and Layer 1 outputs for the two different input patterns will be:

$$\text{for pattern 1: } \mathbf{n}^1 = \begin{bmatrix} 0.8 \\ 0.2 \end{bmatrix},\ \mathbf{n}^2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix},$$

$$\text{for pattern 2: } \mathbf{n}^1 = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix},\ \mathbf{n}^2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$



E18.10 Repeat Exercise E18.9, but use the Hebb rule with decay, Eq. (18.24), instead of the instar learning of Eq. (18.25). Explain the differences between the two responses.