Chapter 10 Associative Memory Networks

We are in for a surprise though!

If we return to our calculated W based on the stored ξ and use $x = (-1,1,-1,-1,-1,1)^T$ as the input then we get the output $y = (-1,1,-1,-1,-1,1)^T$, i.e. the same output as input!

But the x used as input was not stored, then why was it retrieved?

We notice that the x used is identical to ξ, except that the sign of each element has been changed. It is generally true that when we store a vector ξ in a weight matrix W, then we also store its very opposite. You can easily check this with the general relationship on page 100.

But there is even more surprise in store for us. When we store several vectors, then we also store certain combinations of these vectors and some other vectors whose relationship to the fundamental memories is not obvious.

All these “inadvertently” stored vectors are called “spurious attractors”.

Let us look at some spurious attractors.

Chapter 10 Associative Memory Networks

An impressive assortment of spurious attractors generated by learning! (From Haykin, page 700)

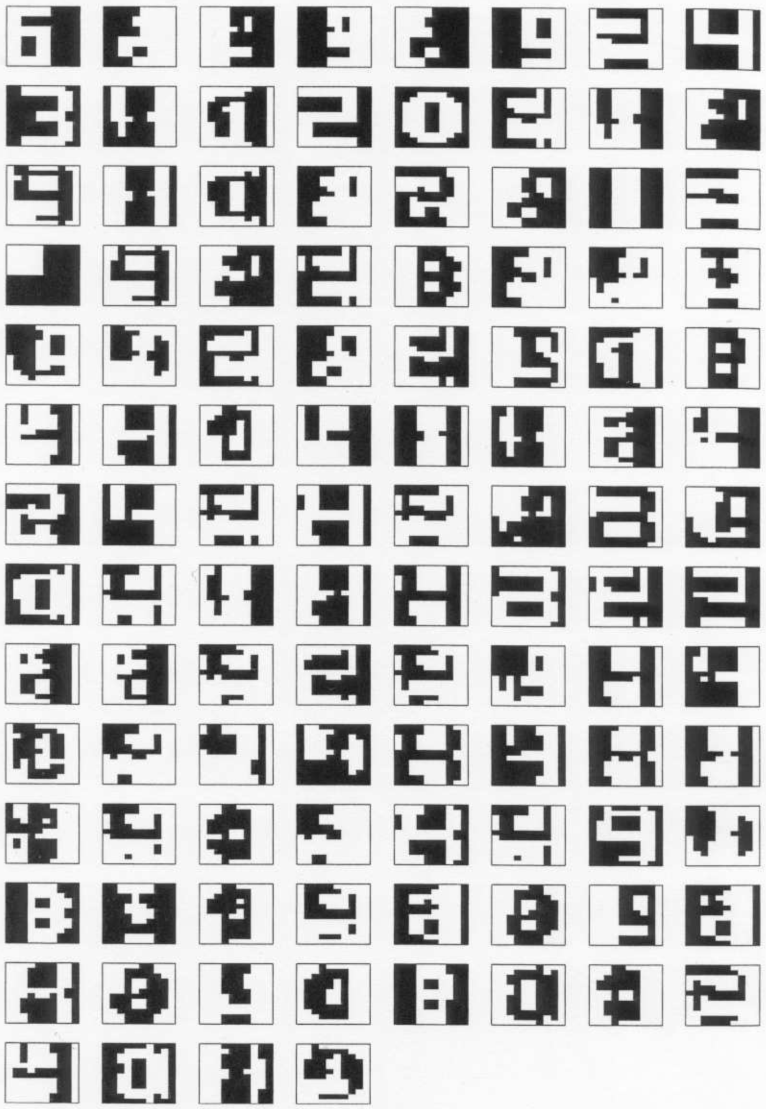


FIGURE 14.20 Compilation of the spurious states produced in the computer experiment

Chapter 10 Associative Memory Networks

Then what does all this mean?

It means that you cannot store memories that are similar to each other, because if you have a slightly corrupted version of one of two similar memories, then you can easily end up in the other one.

It also means that even if the stored memories are not similar to each other, there will be other, spurious memories “in-between”. And if your corrupted initial input is closer to its own fundamental memory than to all the other fundamental memories it is still not certain that the proper fundamental memory will be retrieved, it might well be a spurious attractor instead.

The risk of retrieving the opposite of a fundamental memory is usually not great – your initial input has to be very corrupted for that to happen.

All this means that the associative memory networks as we have described them are far from ideal from a legal witness point of view. But their shortcomings are not unheard of from human experience. So their shortcomings don’t rule them out as first-order models of human memory.

Chapter 10 Associative Memory Networks

The energy landscape

There is an “energy” associated with the states of an associative memory network. It is called energy because Hopfield, who used the energy concept to describe the retrieval process in an associative memory network (in the beginning of the 1980’s) is a physicist and saw the purely formal similarity with energy functions in mechanics.

Each attractor is a minimum, i.e. a lower point than its immediate surroundings, in this energy landscape. If the retrieval process starts from a corrupted memory, then it starts at a high energy and, like a ball in a real landscape, it rolls down to a minimum, hopefully to the right one.

A problem is that the ball rolls in a high-dimensional landscape, which makes it difficult to illustrate on paper.

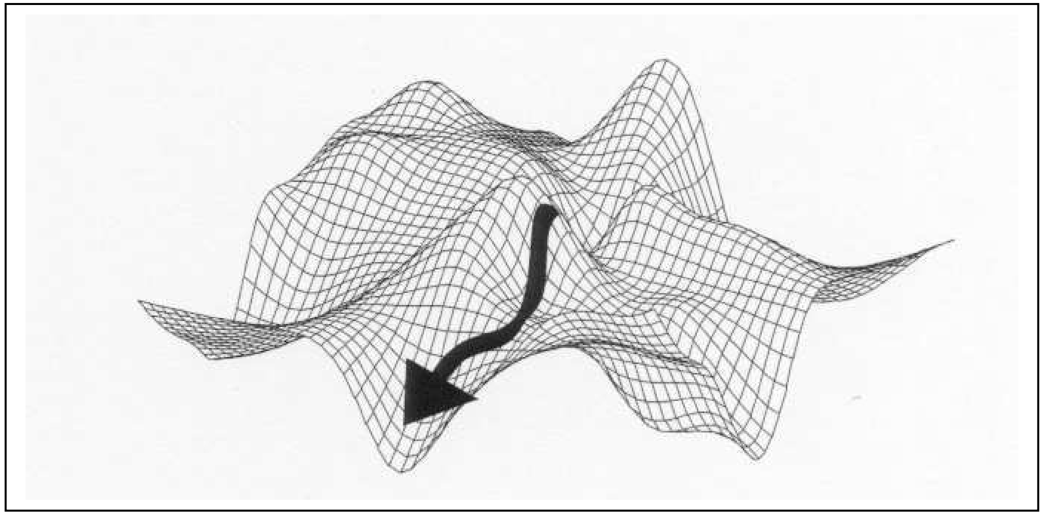
The energy of the spurious attractors is generally higher than the energy of the fundamental memories, so if you can “feel this energy” then you have a chance to say that what you seem to remember might be wrong. The opposite attractors of the fundamental memories have the same low energies as the attractors themselves so in this case you are left without assistance.

Chapter 10 Associative Memory Networks, Energy

The energy of a particular state *x* is defined as

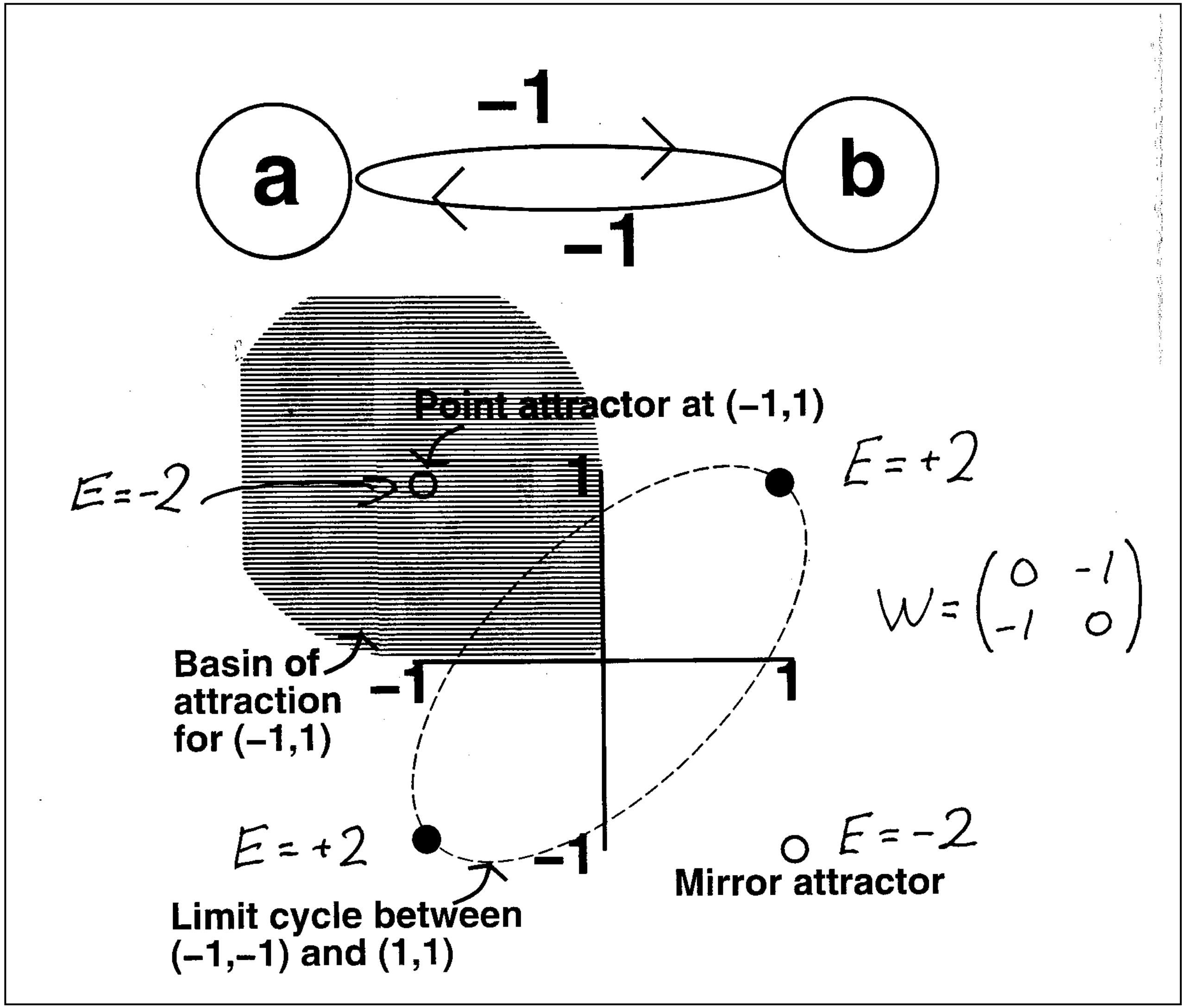
$$E = -\frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{j=1 \\ i\neq j}} w_{ji}x_i x_j$$

The minus sign ensures that we have minima for the “ball to roll into”, rather than peaks to climb. Below is an imaginary energy landscape, from Lytton.



Chapter 10 Associative Memory Networks, Energy

The four possible states of a two-dimensional memory network are shown. One is a fundamental memory, one is its opposite and two lie on a limit cycle as Lytton sees it. Hopfield who updates one element at a time would not have a limit cycle.


a
b
-1
-1
Point attractor at (-1,1)
E=-2
E=+2
1
-1
1
-1
W=$\begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}$
Basin of attraction for (-1,1)
E=+2
o E=-2
Mirror attractor
Limit cycle between (-1,-1) and (1,1)

Chapter 10 Associative Memory Networks
A higher dimensional example

Let us study an example. We begin by storing one vector as a fundamental memory. The particular choice of vector is not important , let us choose $\xi 1$ = [1 1 1 1 1 1 1 1 1 1]'.

The energy minimum has two minima at E = -90, one for x = [1 1 1 1 1 1 1 1 1 1]'
and one for x = [-1 -1 -1 -1 -1 -1 -1 -1 -1 -1]', just as was stated before.

There are several more energy levels, but the levels are discrete since the elements can only assume the values 1 and –1.

If we store two vectors, $\xi 1$ = [1 1 1 1 1 1 1 1 1 1]' and $\xi 2$ = [1 1 1 1 1 -1 -1 -1 -1 -1]', then as expected we have minima at E = -80 for $\xi 1$ and $\xi 2$ and their “opposites”.

Chapter 10 Associative Memory Networks

A higher dimensional example

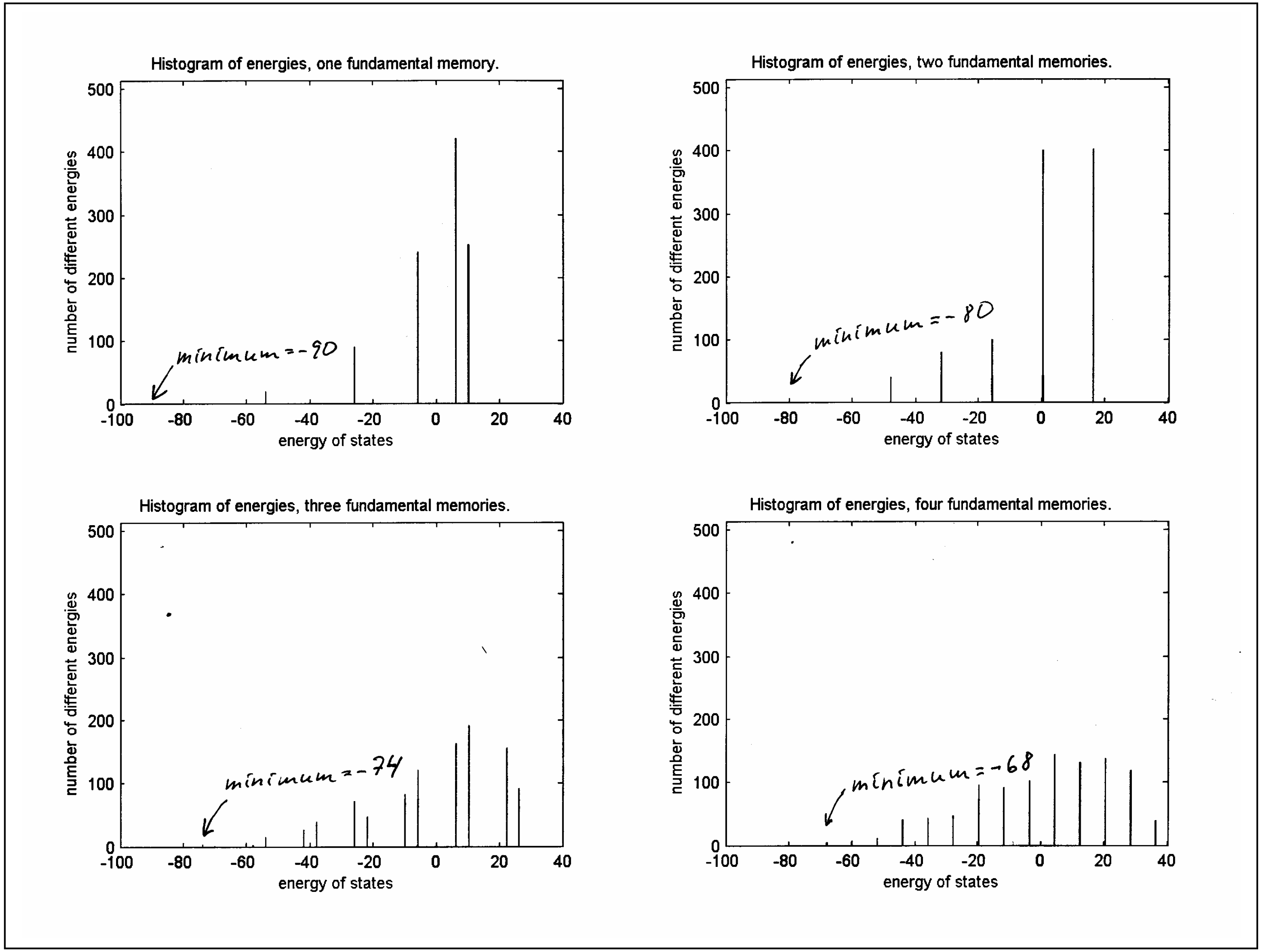
If we store three vectors, $\xi 1$ = [1 1 1 1 1 1 1 1 1 1]' and $\xi 2$ = [1 1 1 1 1 -1 -1 -1 -1 -1]' and $\xi 3$=[1 -1 1 -1 1 -1 1 -1 1 -1]', then we have minima at E = -74 but only at $\xi 2$ and $\xi 3$. The energy at $\xi 1$ is slightly higher, at –70.

If we store four vectors, $\xi 1$ = [1 1 1 1 1 1 1 1 1 1]' and $\xi 2$ = [1 1 1 1 1 -1 -1 -1 -1 -1]', $\xi 3$ = [1 -1 1 -1 1 -1 1 -1 1 -1]' and $\xi 4$ = [1 1 -1 -1 1 1 -1 -1 -1 1]', then we have minima at E = -68 but only at $\xi 2$, $\xi 3$ and $\xi 4$. Again the energy at $\xi 1$ is slightly higher, at –60.

There are more differences between are four cases which become clear when we study the histograms for the energies. Let us do that.

Chapter 10 Associative Memory Networks

A higher dimensional example



Chapter 10 Associative Memory Networks
A higher dimensional example

We see that as the number of fundamental memories increase we get many more intermediate energy levels. Also there is much less difference between the minimum energy levels and other low energy levels. In other words, the energy minima of the attractors are not as distinct when we have more fundamental memories.

We can also expect some spurious attractors to have appeared, i.e. attractors that don't correspond to fundamental memories (others than opposite attractors, which are always present).

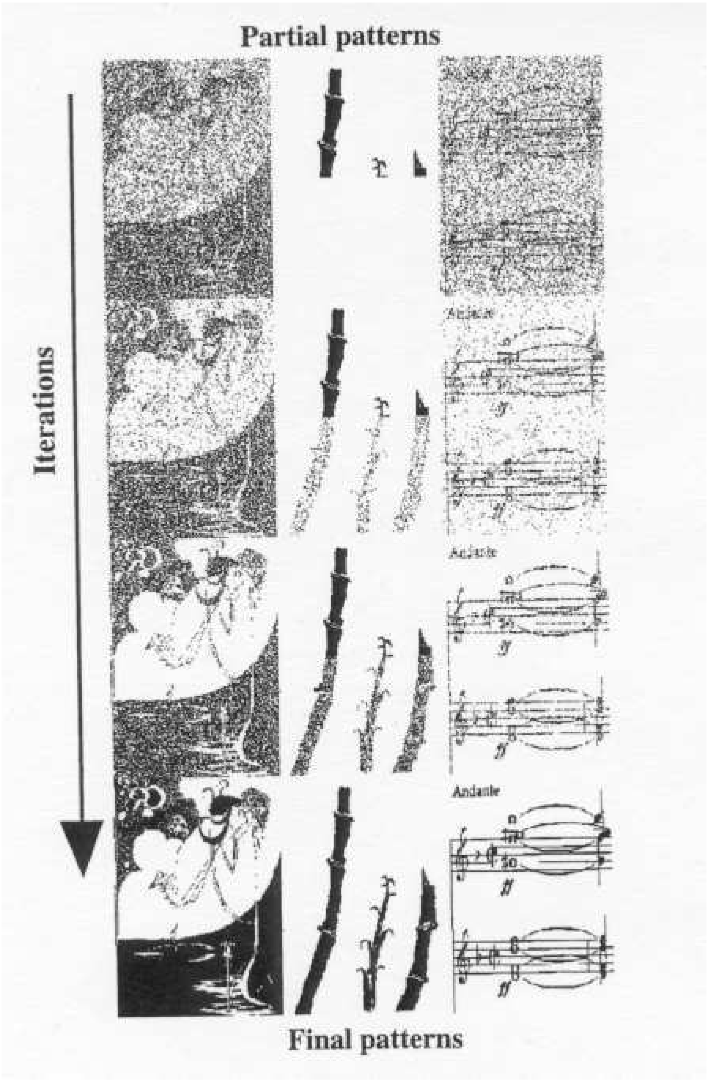
We might wonder if we have lost $\xi 1$ as a fundamental memory, but that is happily not the case. If we let the input to the network be $x = \xi 1$ then the output will be $\xi 1$. This is of course not a big feat, but we haven't lost the fundamental memory.

If we let the input be $\xi 1$ with any one element changed from 1 to –1 we can calculate the energy level at – 36, thus all these slightly corrupted versions of $\xi 1$ lie on a higher energy level than $\xi 1$ itself. We would expect to retrieve $\xi 1$ from such corrupted versions.

Chapter 10 Associative Memory Networks

Retrieval of images in memory

Three very different images are fundamental memories. Corrupted (noisy) versions of the first and third and an incomplete version of the second are keys to start the retrieval process (from Lytton).


Partial patterns
Iterations
Andante
Andante
Andante
Final patterns