Computational Neuroscience: Neuronal Dynamics of Cognition

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

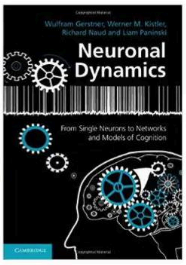
Attractor Networks and Generalizations of the Hopfield model

Wulfram Gerstner

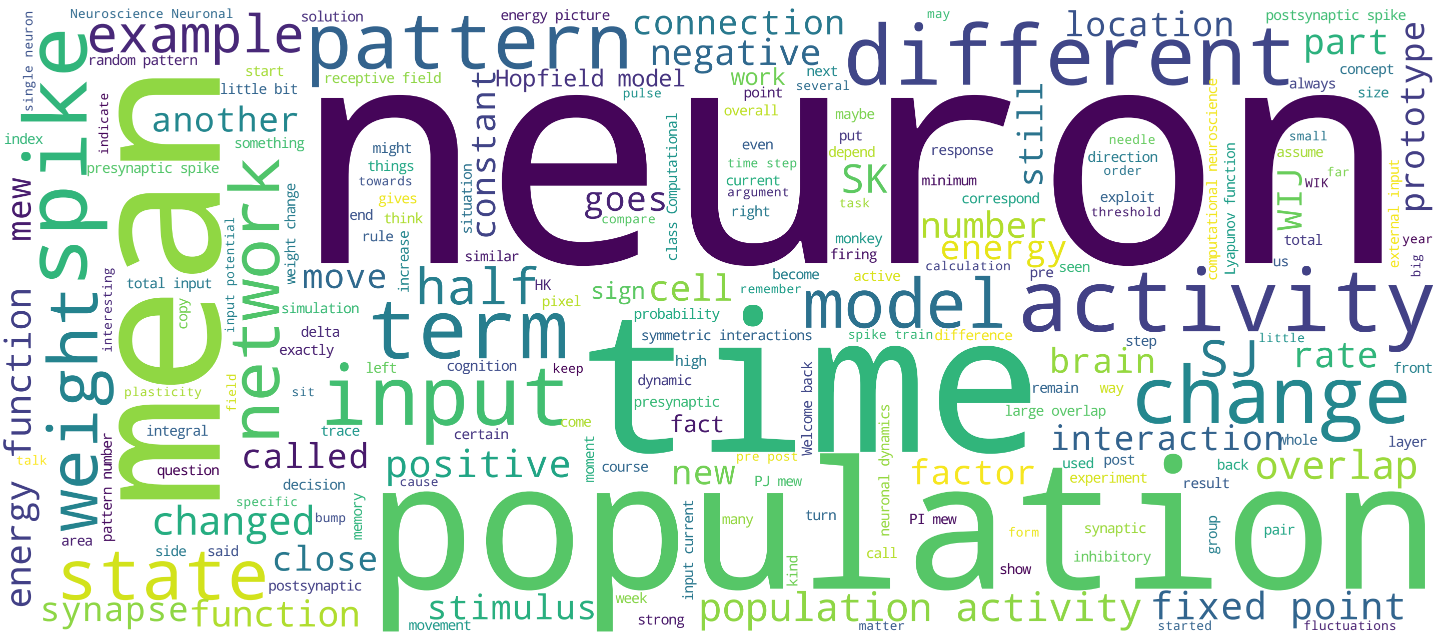
EPFL, Lausanne, Switzerland

Reading:
NEURONAL DYNAMICS
- Ch. 17.2.5 – 17.4

Cambridge Univ. Press



1. Attractor networks
2. Stochastic Hopfield model
3. **Energy landscape**
4. Towards biology (1)
   - low-activity patterns
5. Towards biology (2)
   - spiking neurons

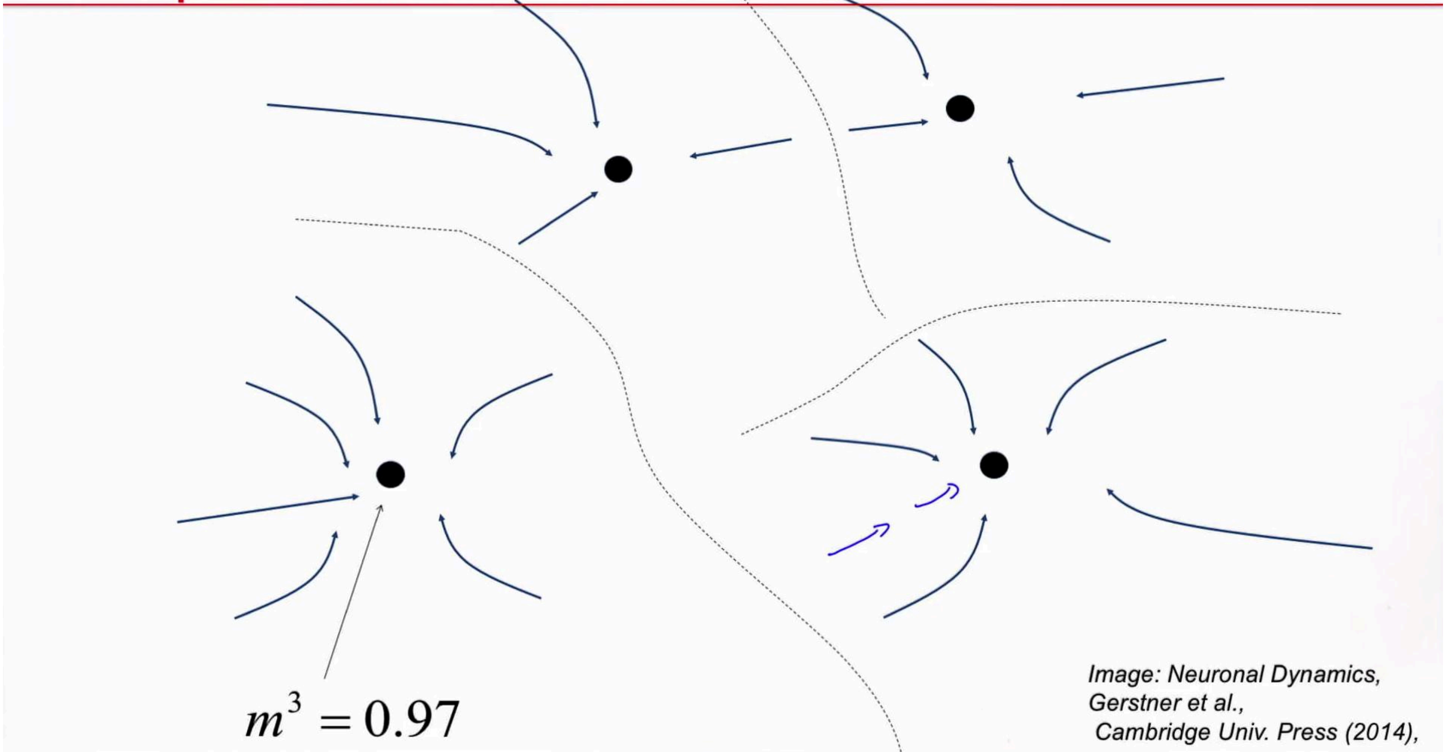


Search MOOC

Video

EPFL

3. Hopfield model = attractor model



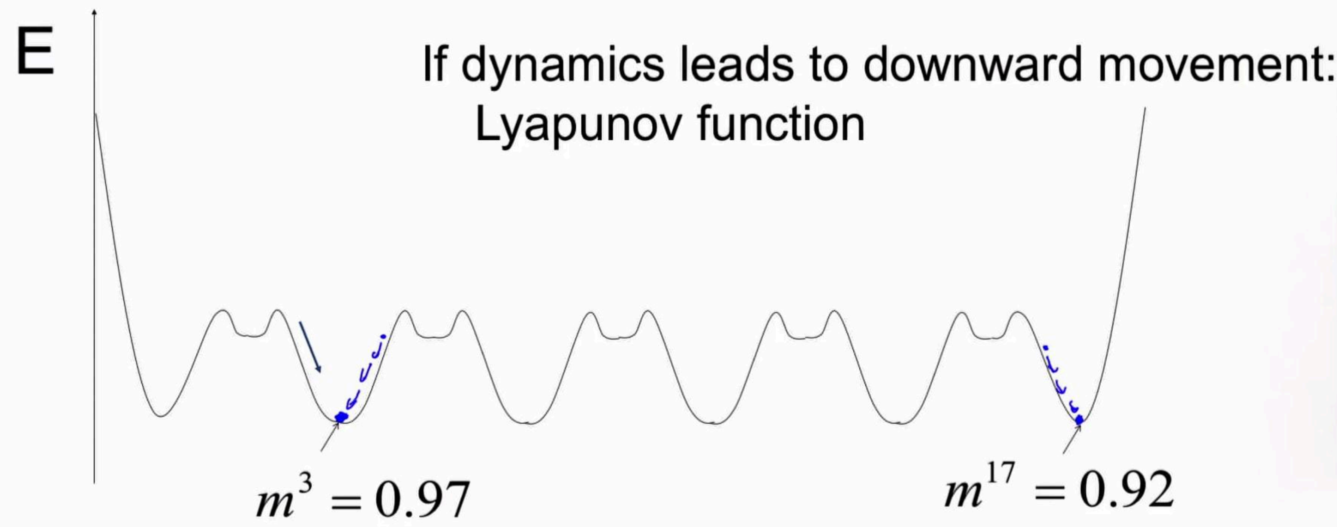
So welcome back to the class competition, neuroscience, and neuronal dynamics of cognition. What you have seen in the previous section is that we can think of the Hopfield model as an attractive model, where over the course of several iterations be to move more and more towards a fixed point. And the final stage would be very close to the fixed point.

Notes

Summary

0m 05s

3. Symmetric interactions: Energy picture

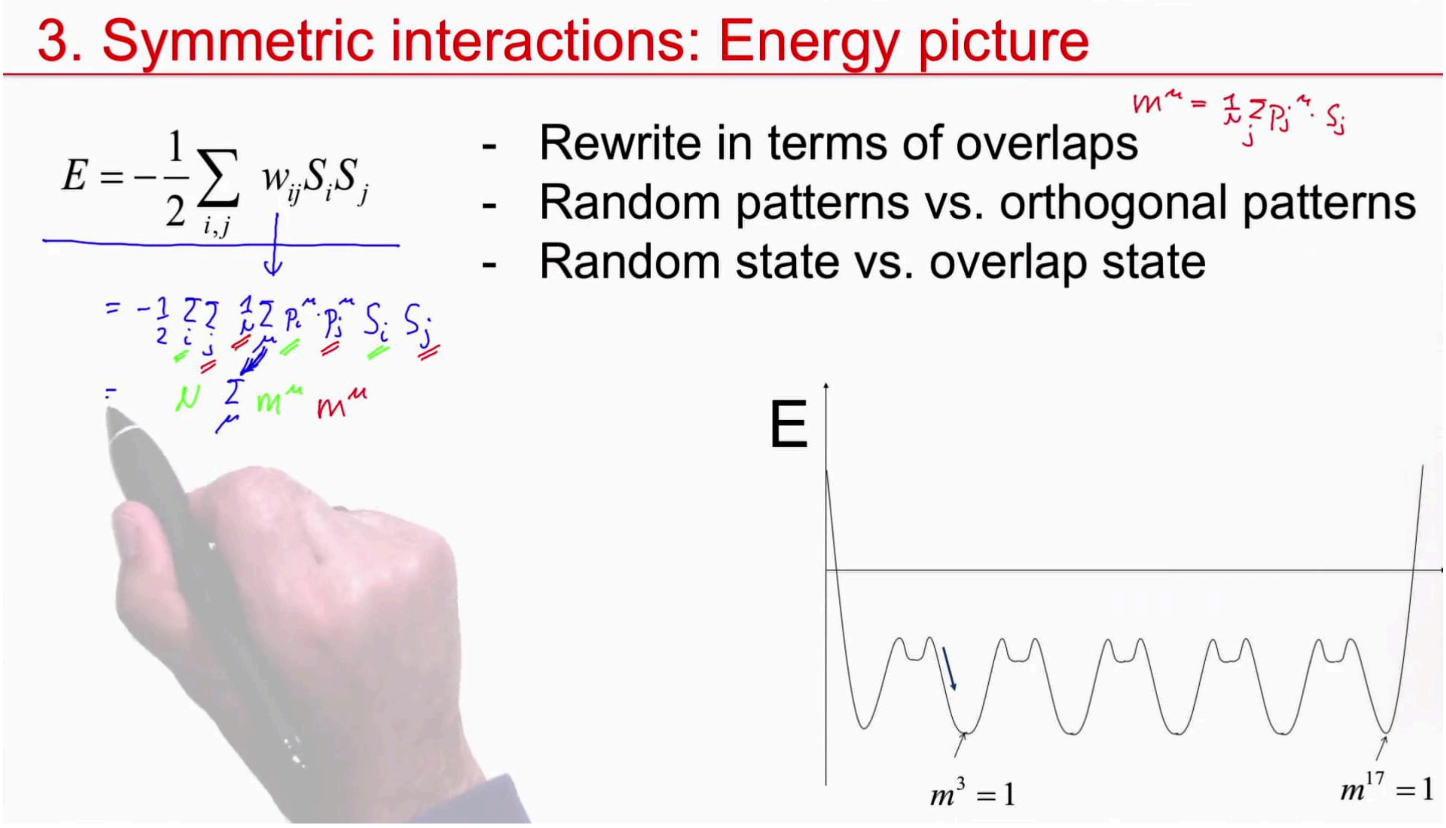


Now, there's an alternative few, which is called the energy model view, you could say, well, my fixed point corresponds to an energy minimum. And then as if I start close to this minimum that starts in one of the valleys, I would move down and approach to minimum, the minimum off the energy function, E corresponds to a state with large overlap, which means that the network has retrieved one of the patterns, for example, pattern three, if I start somewhere else, I would move down and end up in a different attractor, in a different minimum. So the energy concept is mathematical speaking, the concept of Lyapunov function if you have a dynamic, and if you have a function, E, such that, at each step, you've walked downward, or you stay, then you have Lyapunov function.

Notes

Summary

0m 22s



What I propose is following Hopfield that you have an energy function, which is given here is the sum of all neurons, I and all neurons J, so it's a double sum W I J SI SJ. And we will argue in this section, that this energy function Lyapunov function, which means it actually goes down. But in order to see what this energy function is doing, let us first rewrite it in terms of overlaps. So in the case of the Hopfield model, we will use random patterns and have a weight matrix, which is one over N and sum over mew all patterns PI mew PJ mew. This is my weight. And then I have my terms, SI the state of neuron I and SJ, the State of neuron J, and have to sum overall I, I have to sum overall J and that's my factor one half in front. Now, remember, the definition of the overlap, overlap with pattern mew is one over N, sum over PJ mew SJ. Look what we have here the one over N PJ mew SJ sum over J. So this gives to overlap M mew, which means overlap with a pattern number of mew. But now look at this, I have another sum. Sum of I PI mew SI. So I haven't another overlap, except that I missing the factor N. So the factor N remains. And then the only thing that remains is the sum over mew that you still have to minus one half in front.

Notes

Summary

1m 14s

3. Symmetric interactions: Energy picture

$$E = -\frac{1}{2}\sum_{i,j} w_{ij} S_i S_j$$

- Rewrite in terms of overlaps $m^\mu = \frac{1}{N}\sum_j p_j^\mu S_j$
- Random patterns vs. orthogonal patterns
- Random state vs. overlap state

$$= -\frac{1}{2}\sum_i \sum_j \frac{1}{N}\sum_\mu p_i^\mu p_j^\mu S_i S_j$$

$$= -\frac{1}{2} N \sum_\mu m^\mu m^\mu$$

$$= -\frac{1}{2} N \cdot \sum_\mu (m^\mu)^2$$

zero overlap

$\Rightarrow \quad E = 0$

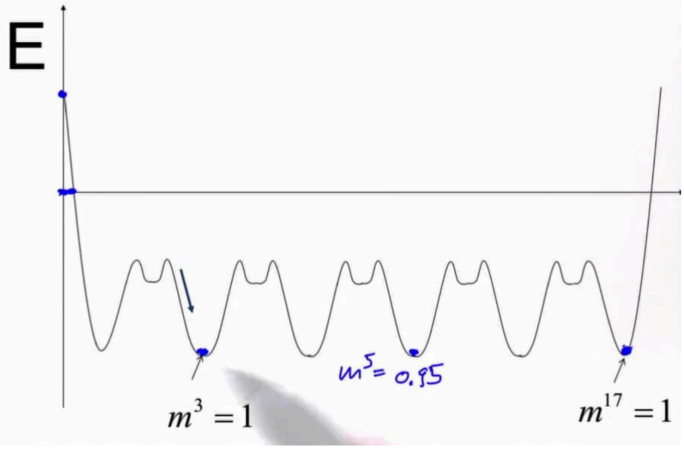
random

$S_i = \pm 1$

$S_j = \pm 1$

$$E = -\frac{1}{2}\sum_i \sum_j w_{ij} S_i S_j$$

$\Rightarrow \quad E \approx 0$



So I have minus one half a factor N and then sum over mew, M mew squared. So far, I considered the case of a large overlap. Now if I have zero overlaps any of the patterns, then the energy will just be zero. More generally just look at the definition of the energy function minus one half, sum of all I sum of all J WIJ SI SJ. And consider an arbitrary state SI is plus-minus one. SJ is plus-minus one, completely random, completely uncorrelated with any of the patterns, that I would have here terms plus-minus one plus-minus one supposed we have arbitrary weights WIJ, that might be positive or negative so I will sum over many terms, plus or minus one. So the energy of a random state will be very close to zero. So there should be random states that have the energy close to zero, then I have if I have the Hopfield wait matrix, I have special states with large overlap might also be stated with which are under correlated with one of the patterns of would have positive energy. Therefore I have a complicated energy landscape. The axis here is completely meaningless, just to indicate that we have an energy landscape with several local minima. And for the Hopfield model, this would correspond to the different retrieval states.

Notes

Summary

3m 04s

3. Symmetric interactions: Energy picture

$$m^\mu = \frac{1}{N}\sum_j p_j^\mu \, S_j$$

$$E = -\frac{1}{2}\sum_{i,j} w_{ij} S_i S_j$$

$$= -\frac{1}{2}\sum_i\sum_j \frac{1}{N}\sum_\mu p_i^\mu p_j^\mu S_i S_j$$

$$= -\frac{1}{2} N \sum_\mu m^\mu m^\mu$$

$$= -\frac{1}{2} N \cdot \sum_\mu (m^\mu)^2$$

zero overlap

$\Rightarrow E = 0$

random

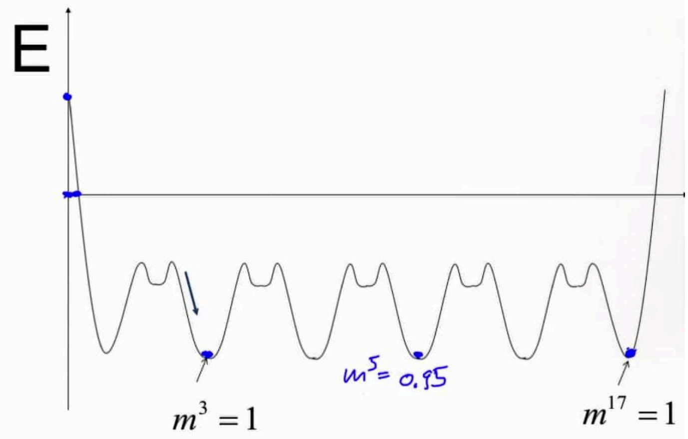
$$E = -\frac{1}{2}\sum_i\sum_j w_{ij} S_i S_j$$

$S_i = \pm 1$

$S_j = \pm 1$

$\Rightarrow E \approx 0$

- Rewrite in terms of overlaps
- Random patterns vs. orthogonal patterns
- Random state vs. overlap state



States with a fairly large overlap with one of the patterns However, I have not shown you that the dynamics would actually lead downward down towards one of the minima.

Notes

Summary

4m 44s

3. Symmetric interactions: Energy/Lyapunov function

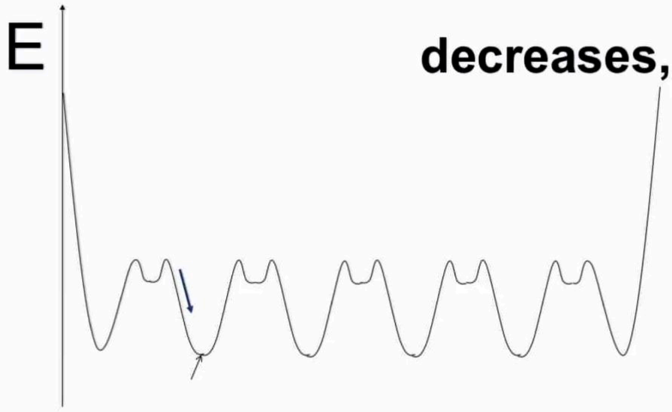
Assume symmetric interaction, $w_{ij} = w_{ji}$

Assume deterministic asynchronous update

$$S_i(t+1) = \mathrm{sgn}[h_i(t)] = \mathrm{sgn}[\sum_j w_{ij} S_j(t)]$$

Claim: the energy $\underline{E = -\frac{1}{2}\sum_{i,j} w_{ij} S_i S_j}$

decreases, if neuron ***k*** **changes**



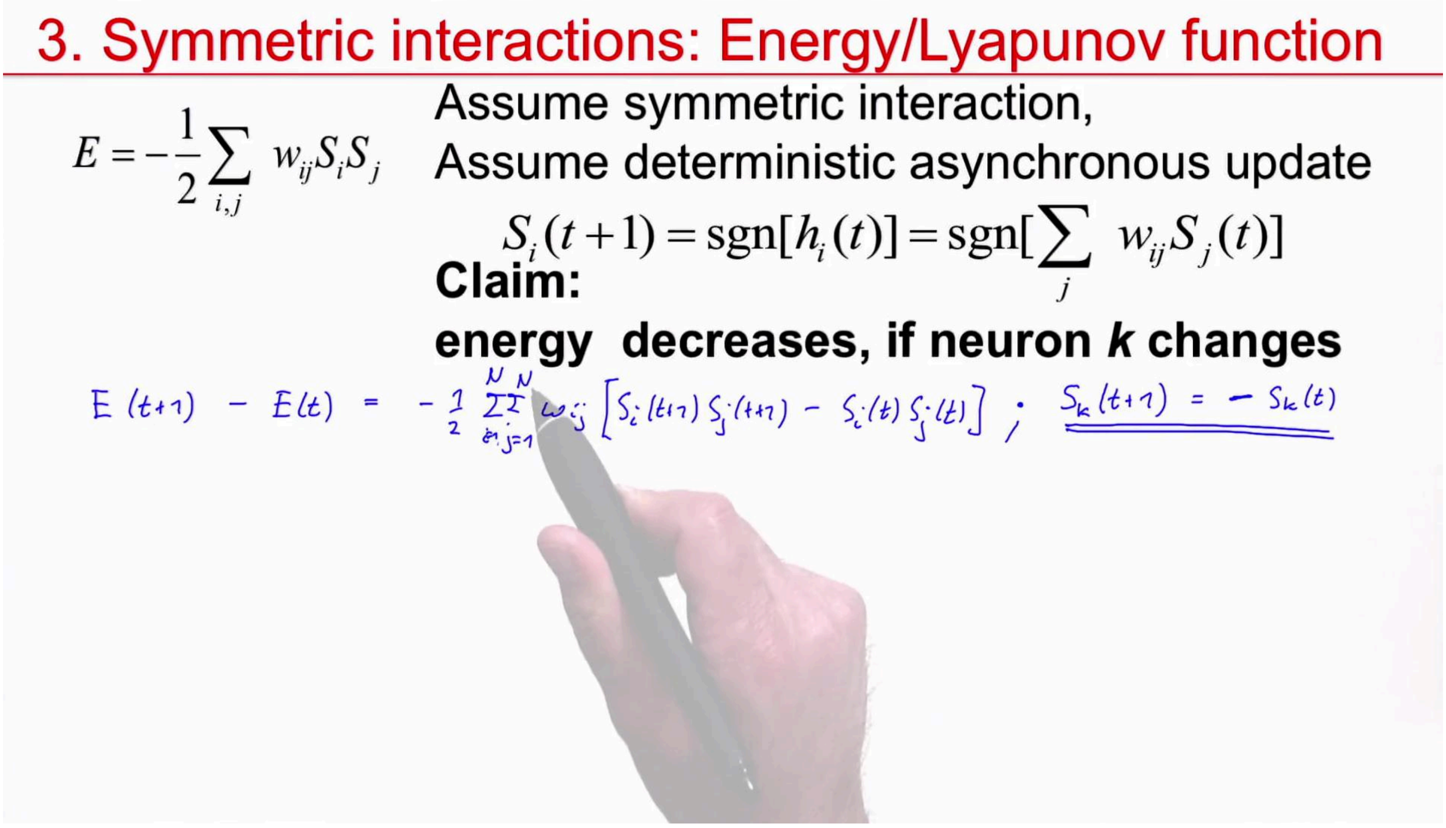
J.J. Hopfield (1982) Neural networks and physical systems with emergent collective computational abilities. Proc. Natl. Acad. Sci. USA 79, pp. 2554–2558

But here we have a nice theorem. Let's define the energy function as before, and let's assume symmetric interactions. That means to weight WIJ, which is from neuron J to neuron I is the same as WJI. Let we assume deterministic asynchronous updates, which means that update one neuron at a time, then the claim is that our energy decreases. If one of the neurons, for example, neuron K changes its state so let's prove this.

Notes

Summary

4m 56s

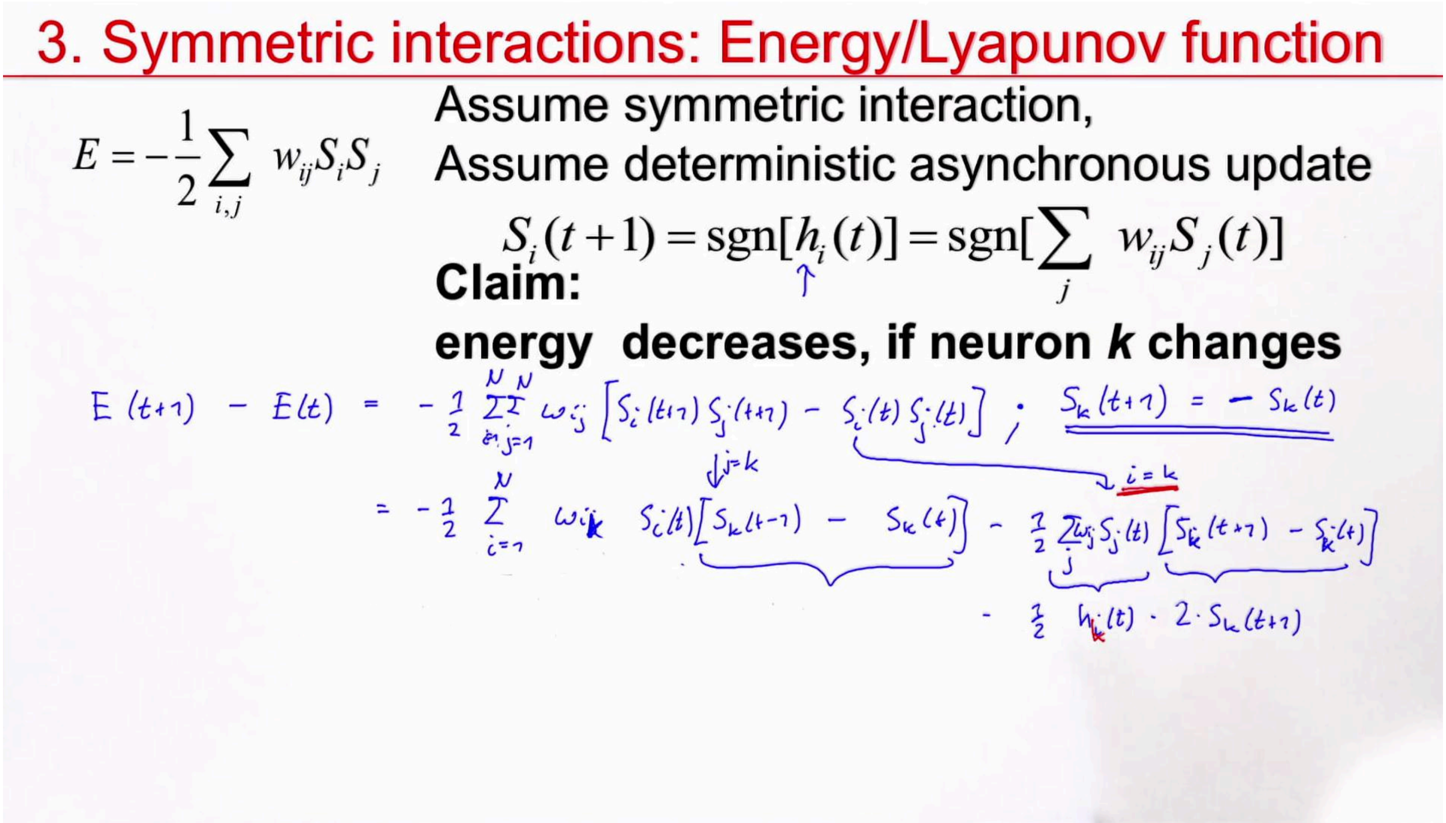


So I started with my energy function. Since we assume asynchronous update, only one neuron in changes, let us consider this neuron. And let's call it K. That means neuron K at time t plus one has changed is different from the value at time t. But these are plus-minus ones though the only way to be different is that is this same with the minus sign. If it was plus one it will become minus one if it was minus one it will become plus one. Now let's look at the energy. And let's compare the energy at time t plus one with the energy at time t. So let's block this in. I have minus one half, sum over I sum over J WIJ SI of t plus one SJ of t plus one minus SI of T SJ of T. Now, only neuron K has changed, all other neurons remain the same. This means that the square brackets will be zero. If neither I nor J is equal to K now two sums here, let me write this explicitly is a sum over I sum over J. So in this sum over i, which runs may be from one to a total of N neurons, say 2000 neurons. And in this sum over J which also runs from one to end. At some point, the J will be equal to K say K is neuron number 25, then there will be the situation when I take this sum the J is equal to 25.

Notes

Summary

5m 31s

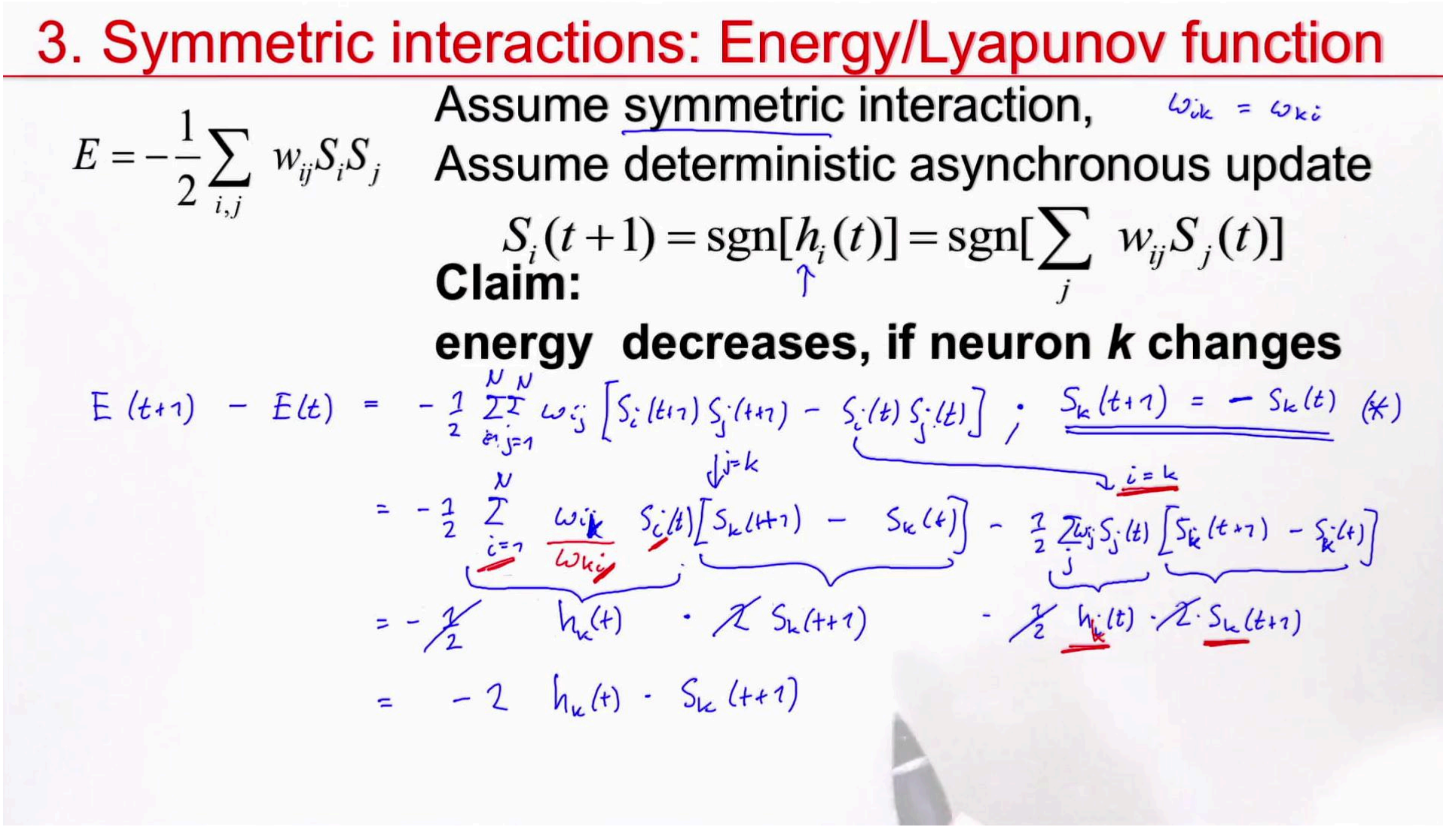


And that's the moment when the neuron changes. So it could be that J is equal to neuron K. J equal to k, and then I have SK of t minus one minus SK of t. And SI is the same SI has not changed SI of T is the same as SI of T plus one WIJ. Now J is now K. So this is WIK. But the sum over i remains if the mind is one half. But it could also be that when I take this sum of I, that this index I take the value 25 that I become equal to k. So this might also be a possibility i equal k in which case I have i fixed. So I has changed I have SI of t plus one minus SI of T. Now SJ J is a free index. But SJ has not changed. Therefore SJT plus one is the same as SJT and as my WIJ which I plugin here. And so this is now a sum over J. Now this is interesting. If you look here because this is just HI of T, this is the total input to neuron i here, I wrote i but I said it's i equal K. So this is really neuron K now neuron K has changed. But if you change this, then SK is just SK at times T it's just the negative of the state at time t plus one. So I can write this as two times SK of t plus one and after factor one i. This was the case i is equal to k. Since i is equal to Ki can actually replace the index i hereby the index key.

Notes

Summary

7m 20s



It's the same year. So now I have HK here and have SK there. So this was the term i equal K. Let's now work on the other term J equal K. Here I have a sum WIKSI. Now here we assume the symmetric interactions we exploit the fact that we have symmetric interactions. And therefore WI K is equal to WKI. So I can replace this term here WIK by WKI. And now you see that I have a sum over i WKISI. But this term is now exactly the field HK of t. Now, look at this difference here. This is the neuron that and that changed. So SK of plus one so this argument should be T plus 1 SK of t plus one minus SK of t is 2 SK of t plus one. That's this relation here. And then I can copy the minus one half. So now we see two things, okay, minus one half, which cuts this two, I've minus one half which cuts this two, I've twice the same term. So I have minus 2 HK of t times SK of t plus one. And now here comes the important inside. Suppose HK is positive. Okay. Neuron k was the one that changed if HK of t was positive, then HSK at t plus one will be plus one. So this product is positive. If H is negative, then sign of HK is minus one and that will be the state it's minus times minus again, it's plus.

Notes

Summary

9m 30s

3. Symmetric interactions: Energy/Lyapunov function

$$E = -\frac{1}{2}\sum_{i,j} w_{ij} S_i S_j$$

Assume symmetric interaction, $w_{ik} = w_{ki}$

Assume deterministic asynchronous update

$$S_i(t+1) = \mathrm{sgn}[h_i(t)] = \mathrm{sgn}[\sum_j w_{ij} S_j(t)]$$

Claim:

energy decreases, if neuron *k* changes

$$E(t+1) - E(t) = -\frac{1}{2}\sum_{i,j=1}^{N\,N} w_{ij}\left[S_i(t+1)S_j(t+1) - S_i(t)S_j(t)\right] \quad ; \quad S_k(t+1) = -S_k(t) \quad (*)$$

$j = k$; $i = k$

$$= -\frac{1}{2}\underbrace{\sum_{i=1}^{N} \cancel{w_{ik}}\, w_{ki}\, S_i(t)}\underbrace{\left[S_k(t+1) - S_k(t)\right]} - \frac{1}{2}\underbrace{\sum_j w_{kj} S_j(t)}\underbrace{\left[S_k(t+1) - S_k(t)\right]}$$

$$= -\frac{\cancel{1}}{\cancel{2}}\, h_k(t) \cdot \cancel{2}\, S_k(t+1) - \frac{\cancel{1}}{\cancel{2}}\, h_k(t) \cdot \cancel{2} \cdot S_k(t+1)$$

$$= -2\, h_k(t) \cdot S_k(t+1) \leq 0$$

✓

So this is always smaller than zero or equal to zero if there is no change. But since we assume that only one neuron unchanged, and this is neuron K, the change is actually negative. And this is what we wanted to show the energy decreases. If neuron k changes, you're done.

Notes

Summary

11m 49s

3. Energy picture

energy picture historically important:
- capacity calculations

J.J. Hopfield (1982) Neural networks and physical systems with emergent collective computational abilities. Proc. Natl. Acad. Sci. USA 79, pp. 2554–2558

D.J. Amit, H. Gutfreund and H. Sompolinsky (1987) Information storage in neural networks with low levels of activity. Phys. Rev. A 35, pp. 2293–2303.

energy picture is a side-track:
- it needs symmetric interactions

energy picture is very general:
- it shows that it should be possible to learn other patterns than mean-zero random patterns

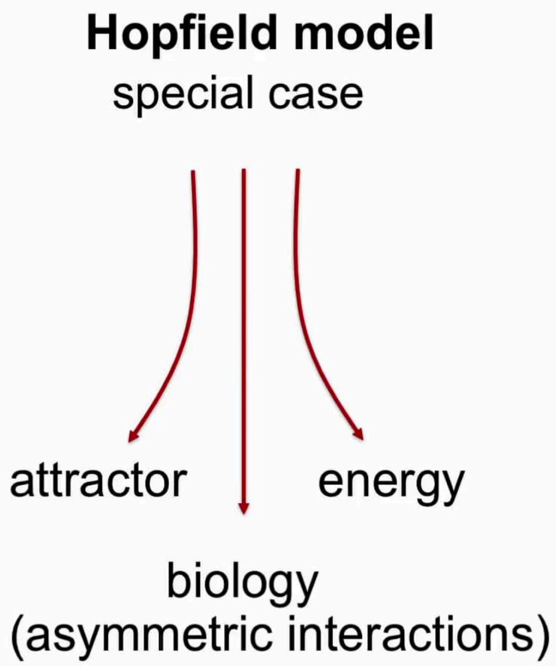
So the energy picture is very useful. And historically, it has been very important. Hopfield wrote his article in 1982, about the neural networks and physical systems with emerging collective properties. And the fact that it was the energy picture, made that many researchers, many physicists joined the field off attractor networks studied these networks. So has been really influential. On the other hand, the energy pictures sort of his sidetrack. It needs symmetric interactions, but we don't have symmetric interactions in biology it has been used for many calculations, but it turns out that some of the calculations you can redo without the energy picture. However, the energy picture suggests that there's something very general interacting neurons in a network can move towards fixed points, even if we don't have random patterns with which exactly have a means zero of the activity.

Notes

Summary

12m 16s

3. Energy picture



So there is something generic going on, and that's what we are going to exploit in the next sections. But before we go there, let's have a look at the quizz.

Notes

Summary

13m 26s

Notes

Summary

13m 36s