
Intelligence

Release 0.0.6

OctoMiao

Oct 29, 2017

CONTENTS

There is a deeply hidden secret, one that killed many, one that saved many. Intelligence, the holy grail of human society, is the ultimate science of all time. As I think about intelligence of many forms, I feel the urge to learn more about them. Hence the notes.

A particular interest is to find resonances in these intelligence. A chinese proverb says, little force can move a huge rock (). Resonances are one of the maneuvers to deflect a huge rock.

CHAPTER
ONE

VOCABULARY

CHAPTER
TWO

QUESTIONS? YES

3.1 Biology

3.1.1 Terms & Concepts

1. cell structure (check figure)

- (a) Nucleus: administration office
- (b) smooth endoplasmic reticulum: high way
- (c) rough endoplasmic reticulum
- (d) Mitochondrion: power plant
- (e) Cytoplasm
- (f) Centriole: combines the surrounding structure to be the centrosome
- (g) Lysosome: trash can, and recycle system
- (h) plasma membrane
- (i) Golgi complex: post office

2. Secretory

- (a) Produced at RER
- (b) Wrapped at Golgi

3. Cell division cycle: For neurons, G0 is when the cells decides to divide or differentiate

- (a) **Interphase**
 - i. G0->G1->S->G2
- (b) Mitosis

4. Cell membrane

(a) **G Protein Coupled Receptors (GPCR): 7 transmembrane**

- i. G proteins have three subunits: alpha beta gamma, which are inactive if no GPCR's are in action
- ii. Receptors for G Protein and important for signaling pathway: ligand binds to GPCR -> GPCR conformational change -> alpha sub-unit exchange its binded GDP for a GTP (now it binds a GTP, and both alpha subunit and beta-gamma complex are active) -> regulate target proteins which are responsible for second messengers -> GTP on alpha subunit then hydrolyze to GDP -> everything comes back together to the original state

5. Protein phosphatase: removes phosphate group from phosphorylated amino acid.
6. Phosphorylation: transfer phosphate group to enzymes
7. Kinase: transfer phosphate group from high-energy to another place
8. Tyrosine kinase: transfer phosphate group to a protein
9. Homeobox: a dna sequence that regulates anatomical development

3.1.2 Techniques

1. **Western blot: separate protein by length of polypeptides**
 - (a) Detergent -> gel electrophoresis -> transfer & incubation
2. Northern blot: RNA
3. **Southern blot: DNA**
 - (a) DNA fragments->gel electrophoresis->ultraviolet
4. **PCR**
 - (a) DNA fragmentation->PCR machine->gel electrophoresis->ultraviolet
5. Clone: somatic cell nucleus + zygote (without nucleus)

3.1.3 Receptors and ligands

1. **Agonist**
 - (a) Has the same effect as a neurotransmitter
2. **Antagonist**
 - (a) Doesn't really work like neurotransmitter but binds to receptor

3.1.4 Neurodevelopment

1. Process: Fertilized egg -> caps -> gastrulation-> neural plate->neural tube-> Anterior-posterior patterning (now we have differences between two ends)-> Dorsal-ventral patterning (differences between dorsal and ventral)
2. Gastrulation: three layers : ectoderm, endoderm, then mesoderm
3. **Polarization of Neurons:**
 - (a) 5 stages: first 2 symmetric, 3 with axon begins, 4 growing in length and branching, 5 morphogenesis and synapse
 - (b) Filopodia:
 - (c) Lamellipodia:
 - (d) Framework: tubulin + actin
4. **ECM: extracellular matrix:**
 - (a) Scaffolding the organ

3.1.5 Refs & Notes

3.2 Visual Cortex

3.2.1 FS in Visual Cortex

Keywords

1. Orientation selective neurons.
2. Feedforward
3. Recurrent

Experiments:

1. Linear amplifications of thalamic input
2. Emergence of pairwise correlations
3. Link between spontaneous and evoked activity. (*RSN (Resting State Networks)*)

Theory:

1. Orientation Selectivity Index: OSI
-

1. Arise of feature-specific (FS) connectivity: induced by learning rule?

3.3 Memory

3.3.1 Time Scale in Memory

Learning and Memory

1. Three categories of memory: [Tetzlaff2012]
 - (a) Working memory: processing information and usually requires with attention.
 - (b) Short-term memory
 - (c) Long-term memory
 2. Evidence of plasticity (Martin et al 2000):
 - (a) Detectability: measured change in synapse when animals is learning or momorizing.
 - (b) Mimicry: replicate the synapse change from one animal to another would result in similar memory or behavior. Not yet experimentally realized.
 - (c) Anterograde alteration: disable synaptic plasticity and learning and memory should stop working.
 - (d) Retrograde alteration: modify synapse weight and the memory should be changed.
-

Why Do We Care about Time Scales

Memory works on many different time scales. This is extremely important for life forms because they need to work out correlations between events in the past and current status.

I would think of permanent memory would serve the purpose of finding out the casualties. However it won't be optimized to work on small time scales. The situation might be quite similar to computers.

Memory Consolidation and Synapse Scaling

Long-term memory requires memory consolidation, which consolidates the synapse locally to enhance memory or transfers memories to other areas.

At the region that connects hippocampus and cortex, LTS (Long-term storage) and STS (Short-term storage) happens at the same time. This indicates that neurons in this cross-section maintain two different types of dynamics, one of which responds to consolidation (i.e., synapse weight change), while the other doesn't.

#TIL# Recall memory can destabilize the corresponding memory itself

In the paper by Tetzlaff [Tetzlaff2013], they mentioned it.

Tetzlaff et al found a solution to this time scale problem [Tetzlaff2013]. In principle, $\frac{dw_{ij}^+}{dt}$ can be Taylor expanded to have all orders of w_{ij}^+ . In this work by Tetzlaff, they included a simple plasticity (zeroth order in Taylor expansion)

$$\mu F_i F_j,$$

as well as a synapse scaling (second order in Taylor expansion)

$$\frac{\mu}{\kappa} (F^T - F_i) (w_{ij}^+)^2.$$

We put those two together,

$$\frac{dw_{ij}^+}{dt} = \mu \left(F_i F_j + \kappa (F^T - F_i) (w_{ij}^+)^2 \right).$$

Such an updating rule for weights contains dynamics of bifurcation. For a region of input frequencies, activity and weights exhibit bistability, thus leading to neurons evolving to different states. Neurons receiving strong inputs have larger weights due to plasticity. Those neurons with large weights evolve into equilibrium points of LTS for the corresponding frequency.

3.4 Alzheimer's Disease

3.4.1 Pathology

1. Amyloid-beta deposition in grey matter for Alzheimer's disease. Accumulation of amyloid-beta then leads to oligomers which finally form plaques. It can **alter intracellular Ca2+ levels**. [LaFerla2008]
2. Amyloid-beta can transmit between humans [Jaunmuktane2015]

3.4.2 Network Dysfunction

1. Hypersynchrony and altered oscillatory rhythmic activity [Palop2016]

3.5 Functional Connectivity

Biological neural networks are quite complicated in the sense maps. Given some input the network generates a output, which is a map of the complicated kind. However, we could also feel the simplicity of such complicated networks. The building blocks of such network are simple neurons which are filters of information.

As a physicist, we have been working with harmonic oscillators. Single oscillators have its frequency of oscillations. Coupled oscillators have their eigenmode, which are hidden collective behaviors of the oscillators. It seems that each oscillator is oscillating chaotically. As we find the eigenmodes, we find the collective patterns. Each oscillator joins some if not all of the collective patterns.

This idea inspires us. If we look at the spiking of each neuron, we might not be able to identify the patterns. However, for coupled neurons, we have to find their collective modes to really SEE the hidden patterns. This idea works for very small networks, the eigenvalues and eigenvectors of which can be found easily. For super large networks, it becomes mission impossible.

In any case, we can persue the idea that the network is composed of basic functional units. There are many levels and aspects of brain connectivity [*BrainConnectivity*].

1. Graph theory: - clustering techniques - community detection, **which is quite interesting**.
2. Morphometric methods
3. Functional brain connectivity: - computing cross-correlations in the time or frequency domain; - mutual information; - spectral coherence.

Community Detection

The article [*BrainConnectivity*] mentioned that **eigenspectrum** is used in community detection.

In social science and complex networks, community detection is a well researched field. A bunch of methods have been developped.

- Q modularity [*Newman2016*];
 - Hyperbolic mapping [*Boguna2010*];
 - Statistical mechanics method such as Gibbs distribution [*Zhang2014*];
 - ...
-

Here are some facts about brain network.

1. It's similar to small world network [*BrainConnectivity*].
2. There are functional hubs [*BrainConnectivity*].
3. Segregation and integration might be key to a functional neural network [*BrainConnectivity*].

3.5.1 References

3.6 Solitary Waves

Solitary waves can be an interesting way of propagating information. In neural networks, solitary waves were found decades ago.

3.6.1 Solitary Waves

We briefly review solitary waves mathematically and physically.

3.7 Signal Decomposition

Acknowledgement

I had some of the ideas when Han Lu was explaining an experiment after one of my lunch. We share some of the ideas.

Signals, electric potential or stimulations, are usually decomposed into different modes to study a system. The most prominent application is the measured signals, but stimulations should also be decomposed to study the response.

Why would we expect the decomposition of signal in a brain to have any significance? **For a neural system of tiny elements to really work out a complicated problem, I expect that some of the neurons have to work collectively in order to solve the problem.** Such collective phenomena might generate observable signals in fMRI etc if the collective neurons are clustering in space. In this case a spherical harmonic decomposition would be able to detect some of the features. Well, this is a little doubtful because the activities might not be global.

3.7.1 Measurement

Application of spherical harmonics to EEG have resulted in interesting experiments [Sivakumar2016].

However, [Sivakumar2016] only used it to decompose the surface of the brain, which is not the complete information. A full map of signal should be a field of 3 dimensional sphere. At each point, we can have information about

1. electric potential,
2. current with direction (flux of charge),
3. electric field with direction.

For electric potential, we can set up a reference point and measure it, which means it's a scalar and we have a number at each point. But for the other two with directions, they are vectors.

The decompositions can be done in different ways, according to the information we want to extract.

1. Globally decompose the signal with spherical harmonics Y at different radius, a multiple expansion for the whole brain; It requires **spin weighted spherical harmonics** etc for vector fields.
2. Locally expand the signal at each point using spherical harmonics; Only need the spherical harmonics.

3.7.2 Stimulation

Stimulations could also be done in a similar way. We apply different modes and check if each mode plays a different role in the brain.

Theoretically, we might think the brain activity might be related to the modes, since each mode stands for different levels of activity in different regions. For example, if we have only $l = 0$ mode at different radius we might not be doing anything, since an isotropic brain is very likely to be a dead one. One of the regions have large activities means we have higher modes excited and since the brain has a fixed anatomy structure, a certain mode indicates activities in different regions.

It's the same for stimulations. Generating a certain mode of electric field or potential will cause a stimulation in different regions.

3.7.3 Theoretical Investigation

We could find the conservation laws of a population in multiple expansion and identify the conserved quantities from population theory. This could be useful for simulations.

3.8 Dispersion Relation

A network of neurons forms a field with the field value being the potential and every other valuable measurable quantities. With this concept set up, we could define plane waves on this network and perform linear stability analysis.

Suppose we have a wave of the form $g(ct - x)$ travelling, which can be Fourier transformed into many plane waves. If the velocity of the wave doesn't depend on the frequency of each Fourier mode, the wave would remain whatever it started as, which is non-dispersive. However, many waves are dispersive and will disperse as they travel. The reason that the waves disperse is that the velocity of each Fourier modes $A(\omega)e^{-i(\omega t - kx)}$ is different such that different Fourier modes will separate out due to the difference in velocity.

Thus we could find out the velocity as a function of frequency to tell how dispersive the waves are, i.e., $c(\omega)$ or $c(k)$, where ω and k are the frequency and wave number respectively.

In fact the phase velocity is defined as $c(\omega) = \frac{\omega}{k}$ while group velocity is $c_g(\omega) = \frac{d\omega}{dk}$. We now established the relations between ω and k , e.g., $D(\omega, k) = 0$. The relation $D(\omega, k) = 0$ is dubbed as **dispersion relation**.

3.8.1 References and Notes

3.9 MISC

3.9.1 Competition between Subnetworks

1. Winnerless Competition: network activity spontaneously switch between same type of subnetworks
2. Cell assemblies: some neurons in mammalian neocortex connected specifically with higher connection probabilities and stronger weights.

3.9.2 RSN (Resting State Networks)

A series of functional networks maintain a high level of coherence even the brain is at rest [Deco2010].

3.9.3 Whole-Brain Computational Connectomics

Whole-brain computational models help neuropsychiatric disorders [Deco2014].

EQUATION SOLVING IN NEUROSCIENCE

Step-by-step demonstration of equation solving in neuroscience, with examples of differential equations from real neuroscience problems.

4.1 Green's Function Method

4.1.1 Green's Function Method

In this section, we demonstrate Green's function method of solving differential equations.

Cable Equation

The cable equation is written as [Gerstner2002]

$$\frac{\partial}{\partial t}u(t, x) = \frac{\partial^2}{\partial x^2}u(t, x) - u(t, x) + i_e(t, x), \quad (4.1)$$

or

$$\frac{\partial}{\partial t}i(t, x) = \frac{\partial^2}{\partial x^2}i(t, x) - i(t, x) + \frac{\partial}{\partial x}i_e(t, x), \quad (4.2)$$

where t, x, i, i_e are all renormalized unit less quantities. For the meaning and definition of them, ref. page 55 of Gerster 2002 [Gerstner2002].

Solutions to Cable Equation

Stationary Solution

Equation (??) can be solved for stationary case, which is

$$\frac{\partial^2}{\partial x^2}u(t, x) - u(t, x) = -i_e(t, x). \quad (4.3)$$

While many methods can be used to solve second order nonhomogeneous differential equations, Green's function is the most general and useful one.

Definition of Green's Function

The idea of Green's function is very simple. TO solve a general solution of equation

$$\frac{d^2}{dx^2}y(x) + y(x) = f(x), \quad (4.4)$$

where $f(x)$ is the source and some given boundary conditions. To save ink we define

$$\hat{L}_x = \frac{d^2}{dx^2} + 1,$$

which takes a function $y(x)$ to $f(x)$, i.e.,

$$\hat{L}_x y(x) = f(x). \quad (4.5)$$

Now we define the Green's function to be the solution of equation (??) but replacing the source with delta function $\delta(x - z)$

$$\hat{L}_x G(x, z) = \delta(z - x).$$

Why do we define this function? The solution to equation (??) is given by

$$y(x) = \int G(x, z) f(z) dz.$$

To verify this conclusion we plug it into the LHS of equation (??)

$$\begin{aligned} & \left(\frac{d^2}{dx^2} + 1 \right) \int G(x, z) f(z) dz \\ &= \int \left[\left(\frac{d^2}{dx^2} + 1 \right) G(x, z) \right] f(z) dz \\ &= \int \delta(z - x) f(z) dz \\ &= f(x), \end{aligned}$$

in which we used one of the properties of Dirac delta distribution

$$\int f(z) \delta(z - x) dz = f(x).$$

Also note that delta function is even, i.e., $\delta(-x) = \delta(x)$.

So all we need to do to find the solution to a standard second differential equation

$$\left(\frac{d^2}{dx^2} + p(x) \frac{d}{dx} + q(x) \right) y(x) = f(x)$$

is do the following.

1. Find the general form of Green's function (GF) for operator for operator \hat{L}_x .
2. Apply boundary condition (BC) to GF. This might be the most tricky part of this method. Any ways, for a BC of the form $y(a) = 0 = y(b)$, we can just choose it to vanish at a and b . Otherwise we can move this step to the end when no intuition is coming to our mind.
3. Continuity at $n - 2$ order of derivatives at point $x = z$, that is

$$G^{(n-2)}(x, z)|_{x < z} = G^{(n-2)}(x, z)|_{x > z}, \quad \text{at } x = z.$$

4. Discontinuity of the first order derivative at $x = z$, i.e.,

$$G^{(n-1)}(x, z)|_{x > z} - G^{(n-1)}(x, z)|_{x < z} = 1, \quad \text{at } x = z.$$

This condition comes from the fact that the integral of Dirac delta distribution is Heaviside step function.

5. Solve the coefficients to get the GF.

6. The solution to an inhomogeneous ODE $y(x) = f(x)$ is given immediately by

$$y(x) = \int G(x, z) f(z) dz.$$

If we haven't done step 2 we know would have some unknown coefficients which can be determined by the BC.

How to Find Green's Function

So we are bound to find Green's function. Solving a nonhomogeneous equation with delta as source is as easy as solving homogeneous equations.

We do this by demonstrating an example differential equation. The problem we are going to solve is

$$\left(\frac{d^2}{dx^2} + \frac{1}{4} \right) y(x) = f(x),$$

with boundary condition

$$y(0) = y(\pi) = 0. \quad (4.6)$$

For simplicity we define

$$\hat{L}_x = \frac{d^2}{dx^2} + \frac{1}{4}.$$

First of all we find the GF associated with

$$\hat{L}_x G(x, z) = \delta(z - x).$$

We just follow the steps.

1. The general solution to

$$\hat{L}_x G(x, z) = 0$$

is given by

$$G(x, z) = \begin{cases} A_1 \cos(x/2) + B_1 \sin(x/2), & x \leq z, \\ A_2 \cos(x/2) + B_2 \sin(x/2), & x \geq z. \end{cases}$$

2. Continuity at $x = z$ for the 0th order derivatives,

$$G(z_-, z) = G(z_+, z),$$

which is exactly

$$A_1 \cos(z/2) + B_1 \sin(z/2) = A_2 \cos(z/2) + B_2 \sin(z/2). \quad (4.7)$$

3. Discontinuity condition at 1st order derivatives,

$$\left. \frac{d}{dx} G(x, z) \right|_{x=z_+} - \left. \frac{d}{dx} G(x, z) \right|_{x=z_-} = 1,$$

which is

$$-\frac{A_2}{2} \sin \frac{z}{2} + \frac{B_2}{2} \cos \frac{z}{2} - \left(-\frac{A_1}{2} \sin \frac{z}{2} + \frac{B_1}{2} \cos \frac{z}{2} \right) = 1 \quad (4.8)$$

Now we combine (??) and (??) to eliminate two degrees of freedom. For example, we can solve out A_1 and B_1 as a function of all other coefficients. Here we have

$$B_1 = \frac{-2/\sin(z/2)}{\tan(z/2) + \cot(z/2)} + B_2,$$

$$A_1 = A_2 + B_2(\tan(z/2) - 1) + \frac{2}{\sin(z/2) + \cot(z/2) \cos(z/2)}.$$

4. Write down the form solution using $y(x) = \int G(x, z)f(z)dz$. Then we still have two unknown free coefficients A_2 and B_2 , which in fact is to be determined by the BC equation (??).
-

The stationary equation (??) can be written as

$$\hat{L}_x u(x) = -i_e(t, x),$$

where $\hat{L}_x = \frac{d^2}{dx^2} - 1$. The boundary condition is the vanishing wave at infinity $u(\pm\infty) = 0$. As we are talking about stationary equation, the source should be time-independent, thus we take only a one dimension Dirac distribution $\delta(x)$ to solve for GF.

The general Green's function is¹

$$G(x, x') = \begin{cases} C_1 e^{-x} + D_1 e^x, & x \leq z, \\ C_2 e^{-x} + D_2 e^x, & x \geq z. \end{cases}$$

Solving Homogeneous Equation

The corresponding homogeneous equation is

$$\frac{d^2}{dx^2} u(x) - u(x) = 0.$$

To find the general solution we assume it has the form

$$u(x) = A e^{\omega x},$$

which is then plugged back into the equation,

$$(\omega^2 - 1)u(x) = 0.$$

We require $\omega^2 - 1 = 0$ to make the solution most general, which leads to

$$\omega = \pm 1.$$

Finally we write down the general solution to this homogeneous equation,

$$u(x) = C e^x + D e^{-x}.$$

In this simple case, BC can be applied to Green's function first², which means

$$G(-\infty, x') = 0,$$

$$G(\infty, x') = 0.$$

¹ In fact this can be obtained by using the Fourier transform method.

² Because the only possibility to make the integral $u(\pm\infty) = \int G(\pm\infty, x')dx' = 0$ satisfy $u(\pm\infty) = 0$ is to make sure GF vanish on the boundaries.

These conditions can significantly simplify the GF,

$$G(x, x') = \begin{cases} D_1 e^x, & x < x', \\ C_2 e^{-x}, & x > x'. \end{cases}$$

Then we use the continuity condition and discontinuity condition,

$$\begin{aligned} G(x'_-, x') - G(x'_+, x') &= 0 \\ \frac{d}{dx} G(x, x') \Big|_{x=x'_+} - \frac{d}{dx} G(x, x') \Big|_{x=x'_-} &= 1, \end{aligned}$$

which is basically

$$\begin{aligned} D_1 e^{x'} - C_2 e^{-x'} &= 0, \\ -C_2 e^{-x'} - D_1 e^{x'} &= 1. \end{aligned}$$

Solving out the coefficients, we get

$$\begin{aligned} D_1 &= \frac{1}{2} e^{-x'}, \\ C_2 &= \frac{1}{2} e^{x'}. \end{aligned}$$

Then we reached the complete and final GF,

$$G(x, x') = \begin{cases} \frac{1}{2} e^{x-x'}, & x < x' \\ \frac{1}{2} e^{x'-x}, & x > x' \end{cases}$$

Given any general source $-i_e(t, x)$, we can write down the solution

$$u(x) = \int G(x, x') (-i_e(t, x')) dx'.$$

As a verification, we integrate out for $i_e(t, x) = 1$,

$$u(x) = - \int_{-\infty}^x \frac{1}{2} e^{x'-x} dx' - \int_x^{\infty} \frac{1}{2} e^{x-x'} dx' = 1,$$

which is exactly the solution given by Mathematica and makes sense.

Physical Meaning

So far we have been dealing with math. What is the actual meaning of GF? To dive into this question we need to review the equation for GF, in this case,

$$\left(\frac{d^2}{dx^2} - 1 \right) u(x) = \delta(x' - x).$$

On the RHS, source term is a delta function, which is just a stimulation to the system at point x' . The textbook shows a graph³ for the case $x' = 0$, where we see the stimulation is given for point $x' = 0$ and the potential drops as we deviate from the stimulated point.

In a stimulation-response system, one of the most important properties is the resonance width, or reaction width, which means the deviation required for the amplitude to drop to $1/e$ of the peak value. In this stationary solution, the

³ Wulfram Gerstner and Werner M. Kistler, Spiking Neuron Models, (2002), Fig. 2.17.

distance is 1 in renormalized unit. To transform back to SI unit, recall that the characteristic length is this problem is $\lambda = \sqrt{\frac{r_T}{r_L}}$.

Just to build a picture, this length is around⁴

$$\lambda = \sqrt{\frac{r_T}{r_L}} = \sqrt{\frac{30\text{k}\Omega \cdot \text{cm}^2 / (2\pi\rho)}{100\text{k}\Omega \cdot \text{cm} / (2\pi\rho)}} = \sqrt{\frac{5 \times 10^{11} \Omega \cdot \mu\text{m}}{3 \times 10^5 \Omega \cdot \mu\text{m}^{-1}}} = 1.2\text{mm}$$

Non-stationary Solution

To solve the most general non-homogeneous cable equation even for non-stationary case, we have to introduce a two-dimensional Dirac distribution $\delta^2(t, x) = \delta(t)\delta(x)$.

Green's function for the most general case should satisfy [Gerstner2002]

$$\frac{\partial}{\partial t}G(t, t'; x, x') - \frac{\partial^2}{\partial x^2}G(t, t'; x, x') + G(t, t'; x, x') = \delta(t' - t)\delta(x' - x).$$

Again to save ink we define

$$\hat{L}_{t,x} = \hat{L}_t - \hat{L}_x,$$

where $\hat{L}_t = \frac{\partial}{\partial t}$ and $\hat{L}_x = \frac{\partial^2}{\partial x^2} - 1$.

The trick is to solve for time dependence first by Fourier transforming the equation to frequency space. To achieve that, we define

$$G(t, t'; x, x') = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} G(t, t'; k, x') e^{ikx} dk. \quad (4.9)$$

On the other hand, Dirac delta is Fourier transformed to

$$\delta(k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \delta(\bar{x}) e^{-ik\bar{x}} d\bar{x} = \frac{1}{\sqrt{2\pi}}, \quad (4.10)$$

which infact gives one of the representations of Dirac delta distribution

$$\delta(\bar{x}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{ik\bar{x}} dk.$$

Applying the transforms of (4.9) and (4.10) to the equation we have

$$\hat{L}_{t,x} \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} G(t, t'; k, x') e^{ikx} dk \right) = \delta(t' - t) \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{ik(x'-x)} dk,$$

which becomes

$$\begin{aligned} & \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{\partial}{\partial t} G(t, t'; k, x') e^{ikx} dk - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} G(t, t'; k, x') \frac{\partial^2}{\partial x^2} e^{ikx} dk \\ & + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} G(t, t'; k, x') e^{ikx} dk = \delta(t' - t) \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{ik(x'-x)} dk, \end{aligned}$$

which is then simplified by removing the integral and common parts

$$\frac{\partial}{\partial t}G(t, t'; k, x') + k^2G(t, t'; k, x') + G(t, t'; k, x') = \delta(t' - t) \frac{1}{\sqrt{2\pi}}.$$

Then we solve this first order differential equation.

⁴ Since opening of ion channels can significantly change the transverse conductivity, this estimation can change significantly in different situations.

4.2 Fourier Transform

4.2.1 Fourier Series and Continuous Fourier Transform

A function $f(x)$ defined on $x \in [-L, L]$ can be decomposed into Fourier series

$$f(x) = \sum_{-\infty}^{\infty} A_n e^{ink_0 x},$$

where

$$k_0 = 2\pi/2L = \pi/L.$$

A continuous Fourier transform is not very different from Fourier series for equation solving. A function $f(x)$ can be written as a convolution of another function $g(x)$,

$$f(x) = \int_{-\infty}^{\infty} g(t) e^{2\pi i x t} dt, \quad (4.11)$$

where $f(x)$ is the Fourier transform of $g(t)$. It is called a **transform** because we have a change of variable from t to x . This form is not convenient for the purpose of equation solving. We usually write it as

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} g(k) e^{ikx} dk,$$

which is derived by plug $k = 2\pi t$ in to Eq. ((??)).

Some Tips about Differential Equations

It's useful to try out Fourier series and Fourier transform on some frequently used operators.

For Fourier series

$$\frac{d}{dx} f(x) = \frac{d}{dx} \sum_{-\infty}^{\infty} A_n e^{ink_0 x} = \sum_{-\infty}^{\infty} A_n \frac{d}{dx} e^{ink_0 x} = \sum_{-\infty}^{\infty} (ink_0) A_n e^{ink_0 x}.$$

As we apply Fourier transform,

$$\frac{d}{dx} f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} g(t) \frac{d}{dx} e^{ixt} dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} g(t) (ixt) e^{ixt} dt,$$

which is more or less similar to the result from Fourier series.

4.2.2 Application in Equation Solving

We provide two examples. The first one is a homogenous equation that is NOT solved using this method. The second equation is an inhomogenous one.

In the first example, we apply Fourier series and Fourier transform to the differential equation of harmonic oscillators,

$$\ddot{x}(t) = -\omega^2 x(t), \quad (4.12)$$

where $\dot{} = \frac{d}{dt}$.

The second example will be a damped oscillator, which is

$$\ddot{x}(t) = -\omega^2 x(t) + f(t). \quad (4.13)$$

Homogeneous Equations

We'll find out that homogenous differential equations doesn't have solutions of the Fourier series type. The reason is that the transform of the solution to a homogeneous equation diverges. However, we still love this example because it shows us that using this method doesn't guarantee that we will find a solution.

Using Fourier Series

Insert Fourier series of $x(t) = \sum_{-\infty}^{\infty} A_n e^{ink_0 t}$ into equation of harmonic oscillator ((?)),

$$\sum_{-\infty}^{\infty} A_n e^{ink_0 t} = -\omega^2 \sum_{-\infty}^{\infty} A_n e^{ink_0 t}.$$

The derivatives will only apply to term $e^{ink_0 t}$, so that the equations becomes,

$$\sum_{-\infty}^{\infty} (ink_0)^2 A_n e^{ink_0 t} = -\omega^2 \sum_{-\infty}^{\infty} A_n e^{ink_0 t}.$$

We move everything to the left side and combine the summations,

$$\sum_{-\infty}^{\infty} ((ink_0)^2 + \omega^2) A_n e^{ink_0 t} = 0.$$

The only possible solution to this equation is that the coefficient is zero, i.e.,

$$(ink_0)^2 + \omega^2 = 0,$$

which has a solution of interest

$$k_0 = \pm\omega/n.$$

This result is the dispersion relation like solution to the system. However, we can not construct the original solution from this relation only. The reason is that the equation is homogeneous.

Using Fourier Transform

We assume that the function $x(t)$ that we are looking for is a Fourier transform of another function $\hat{x}(\tau)$

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{x}(k) e^{ikt} dk.$$

Insert this transform into the equation of harmonic oscillator ((?)),

$$\frac{d^2}{dt^2} \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{x}(k) e^{ikt} dk = -\omega^2 \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{x}(k) e^{ikt} dk,$$

which is simplified to

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} ((ik)^2 + \omega^2) \hat{x}(\tau) e^{ikt} dk.$$

We find ourselves in the same situation as the Fourier series solution. No solution is found.

Inhomogeneous Equations

For an inhomogeneous equation ((?)), we can apply the same trick. However, we use Fourier transform for generality. Please note that the Fourier transform requires the solution to be defined on a range of the argument.

Before we actually work out the case, here are some tricks. We always use hat to denote the inverse Fourier transformed equation. Fourier transform of the equation actually will be a replacement of the following,

$$\begin{aligned} f(t) &\rightarrow \hat{f}(k) \\ x(t) &\rightarrow \hat{x}(k) \\ \dot{x}(t) &\rightarrow (ik)\hat{x}(k) \\ \ddot{x}(t) &\rightarrow (ik)^2\hat{x}(k). \end{aligned}$$

Using these rules, we find out the Fourier transform equation of ((?)),

$$(ik)^2\hat{x}(k) + \omega^2\hat{x}(k) = \hat{f}(k),$$

which has a solution

$$\hat{x}(k) = \frac{\hat{f}(k)}{-k^2 + \omega^2},$$

so that the final solution we are looking for becomes

$$\begin{aligned} x(t) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{x}(k) e^{ikt} dk \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\hat{f}(k)}{-k^2 + \omega^2} e^{ikt} dk. \end{aligned}$$

We know $f(t)$ so $\hat{f}(k)$ can be calculated.

4.2.3 Application to Neuroscience

4.3 Laplace Transform

4.3.1 Laplace Transform

Laplace transform is useful in equation solving. By definition, Laplace transform transforms a function $f(x)$ defined on $x \geq 0$ into another function by calculating the convolution

$$F(s) = \int_0^{\infty} e^{-sx} f(x) dx,$$

or symbolically,

$$F(s) = \mathcal{L}_s(f(x)).$$

A table of important Laplace transforms can be found on mathworld.wolfram.com. Here we steal some of the commonly used.

$$\begin{aligned} \mathcal{L}_s(1) &= \frac{1}{s} \\ \mathcal{L}_s(x^n) &= \frac{n!}{s^{n+1}} \\ \sin(kx) &= \frac{k}{k^2 + s^2} \\ \cos(kx) &= \frac{s}{k^2 + s^2}. \end{aligned}$$

It can also be applied to differentials.

$$\begin{aligned}\mathcal{L}_s\left(\frac{df(x)}{dx}\right) &= s\mathcal{L}_s(f(x)) - f(0) \\ \mathcal{L}_s\left(\frac{d^2f(x)}{dx^2}\right) &= s^2\mathcal{L}_s(f(x)) - sf(0) - \left.\frac{df(x)}{dx}\right|_{x=0}.\end{aligned}$$

Laplace Transform of Differentials

The general form is

$$\mathcal{L}_s(f^{(n)}(x)) = s^n\mathcal{L}_s(f(x)) - \sum_{m=1}^n s^{n-m}f^{(m-1)}(0).$$

Integrals with upper limit as argument also transform nicely.

$$\mathcal{L}_s\left(\int_0^x f(x')dx'\right) = \frac{\mathcal{L}_s(f)}{s}.$$

Laplace transform is useful in solving differential equations because it transforms many equations into fractions and polynomials. A simple example is the harmonic oscillators. The equation of motion is

$$\ddot{x}(t) = -\omega^2 x(t),$$

which is transformed into

$$s^2X(s) - sx(0) - \dot{x}(0) = -\omega^2 X(s).$$

The solution to it is

$$X(s) = \frac{sx(0) + \dot{x}(0)}{s^2 + \omega^2} = x(0)\frac{s}{s^2 + \omega^2} + \frac{\dot{x}(0)}{\omega}\frac{\omega}{s^2 + \omega^2}.$$

We can spot sin and cos from the solution, or more generally perform an inverse Laplace transform,

$$x(t) = x(0)\cos(\omega t) + \frac{\dot{x}(0)}{\omega}\sin(\omega t).$$

This example is too simple sometimes naive. However, it shows the spirit.

Caveats

Laplace transform has a lot of counter intuitive expressions.

1. Laplace transform of product of two functions $f(x)g(x)$ is **NOT** the product of the Laplace transforms $\mathcal{L}_s(f)\mathcal{L}_s(g)$. However, if one of the functions is a constant, say $f(x) = 5$, we can prove that the Laplace transform of $5g(x)$ is $5\mathcal{L}_s(g)$.
2. The product of two Laplace transforms $\mathcal{L}_s(f)\mathcal{L}_s(g)$ is the Laplace transform of a convolution

$$\int_0^x f(x')g(x-x')dx'.$$

3. Small s corresponds to large x , due to the nature of the exponential suppression in Laplace transform. For example, for small argument x , the function e^{kt} becomes almost 1, meanwhile, the Laplace transform of the function $1/(s-k)$ becomes $1/s$ under large s . We can see that the two limits are constant since the Laplace transform of 1 is $1/s$.

4. In so many circumstance the Laplace transform doesn't exist simple because the integral doesn't converge. Please beware of this and use the transform only when it exists.

Inverse Laplace Transform

We do not usually use the general form of inverse Laplace transform since we can find it in the table. Nevertheless we write it down here.

$$f(x) = \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} F(s) e^{sx} ds.$$

By defining $s = is^\dagger$, we can rewrite the formula

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(is^\dagger) e^{is^\dagger x} ds^\dagger.$$

The more interesting application is to solve matrix differential equations. For any equations

$$\partial_x \mathbf{f}(x) = \mathbf{A} \mathbf{f}(x),$$

Laplace transform takes it to the form

$$s\mathbf{F}(x) - \mathbf{f}(0) = \mathbf{A}\mathbf{F}(s).$$

The solution is

$$\mathbf{F}(x) = \frac{1}{s\mathbf{I} - \mathbf{A}} \mathbf{f}(0).$$

So the final solution for $\mathbf{f}(x)$ is

$$\mathbf{f}(x) = \mathcal{L}^{-1} \left(\frac{1}{s\mathbf{I} - \mathbf{A}} \right) \mathbf{f}(0).$$

We could work out the Taylor expansion of solution,

$$\frac{1}{s\mathbf{I} - \mathbf{A}} = \frac{\mathbf{I}}{s} + \frac{\mathbf{A}}{s^2} + \frac{\mathbf{A}^2}{s^3} \dots$$

The inverse Laplace transform can be done simply term by term,

$$\mathcal{L}^{-1} \left(\frac{1}{s\mathbf{I} - \mathbf{A}} \right) = \mathbf{I} + x\mathbf{A} + \frac{1}{2!} (x\mathbf{A})^2 + \dots = e^{x\mathbf{A}}.$$

Finally we obtain the formal solution of the system, which is

$$\mathbf{f}(x) = \exp(x\mathbf{A}) \mathbf{f}(0).$$

Only Works for Constant Coefficients

This result only works for constant coefficients. In general, if the matrix A depends on the argument x , the solution can be systematically calculated using the so called [Magnus Expansion](#). However, it is as tedious as a numerical solution.

4.3.2 Application to Neuroscience

RANDOM WALKS

I am learning about random walks.

Here is a set lectures by Sid Redner.

5.1 Random Walks

There exists several simple models of random walks. It can be either fixed stepsize but random directions or varying stepsize.

5.1.1 Fixed Stepsize

A simple model is that we assume each step is a step of size ϵ , but with random directions. The position of this random walker at step N is the summation of all the steps (vectors),

$$\vec{X} = \sum_i^N x_i,$$

where x_i is the vector that represents step i.

What we want to find out is the places that the random walker explored after N steps. The corresponding quantity that represents it is $\sqrt{\langle \vec{X}^2 \rangle}$.

From the idea of random walk, we know that

$$\begin{aligned}\langle \vec{x}_i \rangle &= 0 \\ \langle \vec{x}_i \cdot \vec{x}_j \rangle &= 0.\end{aligned}$$

Thus

$$\begin{aligned}\langle \vec{X}^2 \rangle &= \sum_i \vec{x}_i \cdot \vec{x}_i \\ &= N\epsilon^2.\end{aligned}$$

Then we find out that

$$\bar{X} = \sqrt{\langle \vec{X}^2 \rangle} = \sqrt{N}\epsilon. \tag{5.1}$$

5.1.2 Significance of Dimension

From the root-mean-squared distance Eq. (??) we can define the density of points. Suppose we have a continuous version of this random walk. After time t , the random walker walked a distance \sqrt{t} . Meanwhile the random walker explored a region of radius \sqrt{t} , which corresponds to a volume $V \propto \sqrt{t}^d$. The density of walked points is defined as

$$\rho \propto \frac{t}{\sqrt{t}^d} = t^{1-d/2},$$

where d is the dimension of the space.

We spot this critical dimension $d = 2$.

1. $d < 2$: the density of points at $t \rightarrow \infty$ becomes $\rho \rightarrow \infty$. This called recurrent behavior. We are sure that after infinite time, we are going back to a point that we visited before.
2. $d = 2$: the density of points at $t \rightarrow \infty$ becomes $\rho \rightarrow \text{Constant}$. **This derivation is wrong about this critical case.** It should be $\rho \rightarrow \ln t$
3. $d > 2$: the density of points at $t \rightarrow \infty$ becomes $\rho \rightarrow 0$. This is called transit behavior. We are not sure that we could go back to a point that we visited before.

5.1.3 Refs & Notes

1. Sid Redner's Lectures @ Santa Fe Institute

MACHINE INTELLIGENCE

6.1 Artificial Neural Network

6.1.1 Basics

Artificial neural network can be used as a universal approximator. We all had the experience of guess the solution to some particular problems. Artificial neural network serves as a general one that can be used to approximate a lot of functions.

The simplest example is that we can always describe a function $f(x)$ as

$$y(x) = \sum_{k=1}^N v_k * \text{activation}(w_k * x + u_k),$$

where **activation** is a function that is related to the property of single neurons. Precisely speaking, this decomposition of the function is a network composed of N neurons.

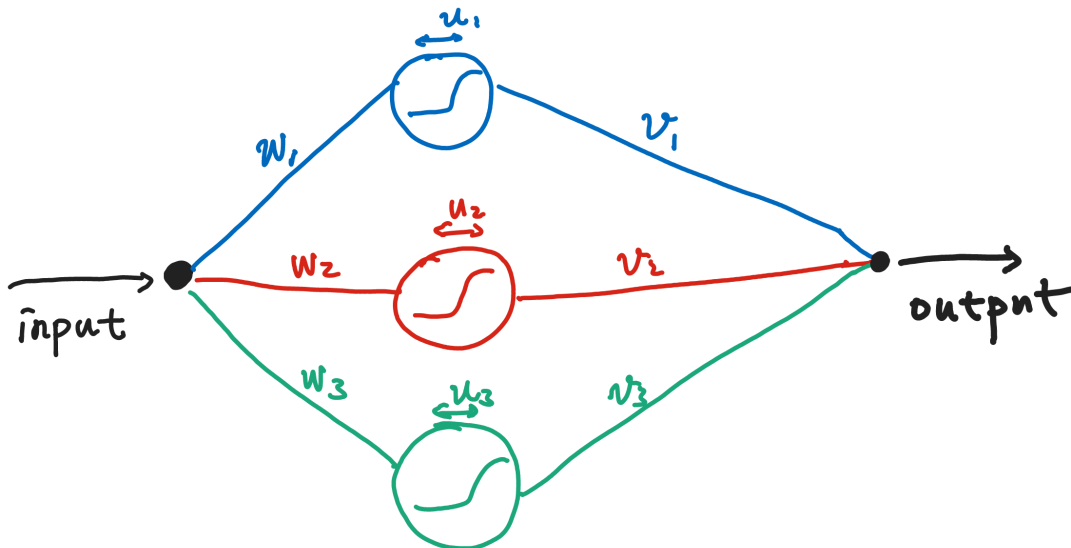


Fig. 6.1: A very simple artificial neural network.

How exactly does it work?

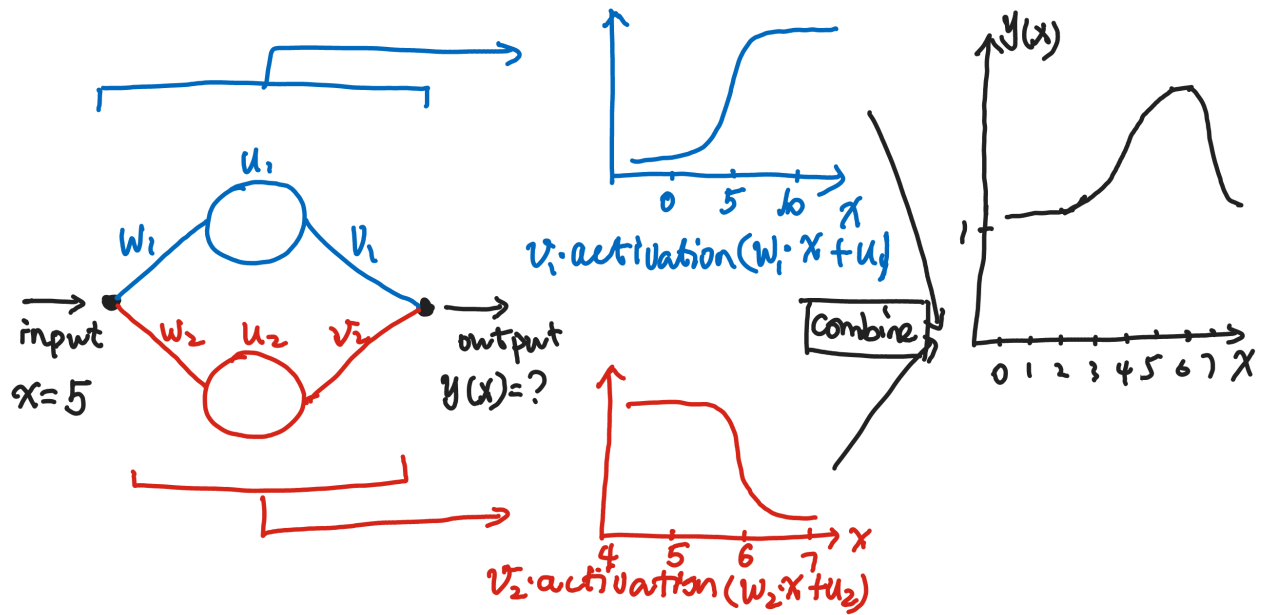


Fig. 6.2: Illustration of how ann works as an approximator.

The only problem, however, is that we need to find out the value of the parameters, thus training. For training we either need some real data that should be approximated by this network or a conservation law.

6.1.2 Solving Differential Equations

For a differential equation, we have a natural conservation law. For example, equation

$$\frac{d}{dt}y(t) = -y(t)$$

means that the quantity

$$\frac{d}{dt}y(t) + y(t)$$

is conserved and is exactly 0. If we ever try to use ANN to approximate the function $y(t)$, this conservation law should be satisfied and it's the only law that the approximator should obey.

Using the network, we know that for each argument t_i , we should have an output

$$y_i = 1 + t_i v_k f(t_i w_k + u_k).$$

By training, we are talking about minimization the deviation of the quantity from the conserved value. We devise a function that describes the deviation and name it the cost,

$$I = \sum_i \left(\frac{dy_i}{dt} + y_i \right)^2.$$

We could actually calculate dy/dt using the approximator,

$$\frac{dy}{dt} = v_k f(t w_k + u_k) + t v_k f(t w_k + u_k) (1 - f(t w_k + u_k)) w_k,$$

where we denote the **activation** function f . Then we can parameterize the cost function using the parameters

$$I = \sum_i (v_k f(tw_k + u_k) + tv_k f(tw_k + u_k)(1 - f(tw_k + u_k))w_k + y)^2.$$

The final step is to find the parameters so that this cost is minimized.

Activation function or trigger function

One of the useful activation function is

$$f(x) = \frac{1}{1 + \exp(-x)}.$$

6.2 Neural Networks and Finite Element Method

In some sense, feed forward neural network is similar to finite element method.

6.3 Boltzmann Machine

Boltzmann machine is much like a spin glass model in physics. In short words, Boltzmann machine is a machine that has nodes that can take values, and the nodes are connected through some weight. It is just like any other neural nets but with complications and theoretical implications.

6.3.1 Boltzmann Machine and Physics

To obtain a good understanding of Boltzmann machine for a physicist, we begin with Ising model. We construct a system of neurons $\{s_i\}$ which can take values of 1 or -1, where each pair of them s_i and s_j is connected by weight J_{ij} .

This is described as a Boltzmann machine, or spin glass in physics. Spin glass is a type of material that is a composite of many spins pointing in different directions. In principle spin glass is hard to calculate.

Nevertheless we can make simplifications to this model. We require each spin to be connected to its nearest neighbours only. Such a model is called Ising model.

Intuitively, those spins can be viewed as tiny magnets that can point up or down only. Each spin interacts with its neighbours. These interactions are calculated in terms of energy,

$$E = - \sum_{i,j} J_{ij} s_i s_j.$$

Why do we care about energy? For a physics system, low energy means stable while high energy means unstable since it might automatically change its configuration into low energy state. That being said, a system of spins is stable if the energy of all the interactions is low.

To find out a low energy state, one of the numerical methods is Monte Carlo method.

States

We have been talking about the word state without being specifying the definition of it. In fact we can think of two different pictures of states. For the purpose of this discussion, we consider a system of N particles and each of the particle has m degrees of freedom.

The first strategy is to set up a $N \times m$ dimension space and describe the state of the whole system with on point in such a space. The distribution of the points can be determined by the corresponding categories of distribution functions. This is dubbed as Γ space.

The second strategy is to use a space of m dimensions where each particle of the system is a point in such a space. Such a space is called μ space. In μ space, the distribution of each particle state is calculated using BBGKY chain.

Once the macroscopic properties of the system is assigned, the all possible states that leads to this macroscopic state show up with equal probability, aka, principle of **equal a priori probabilities**.

Partition Function

Partition function Z is useful as we calculate the statistical properties of the network,

$$Z = \sum e^{-E}.$$

With partition function defined, the distribution of states is

$$P = \frac{1}{Z} e^{-E}$$

6.3.2 Application of Boltzmann in Learning

In many cases, we would like the machine to learn about the pattern of input. Probabilistically speaking, we are working out the probability distribution of the input data using a Boltzmann machine,

$$\Delta p = p_{\text{Model}} - p_{\text{Data}}.$$

Or for reasons of log-likelihood, we use the ration of them

$$\Delta \log p = \log p_{\text{Model}} - \log p_{\text{Data}}.$$

In terms of Boltzmann machine weights, which are to be determined, the log probability is proportional to the energy

$$\Delta \log p \sim \sum_{ij} w_{ij} s_i s_j |_{\text{Model}} - \sum_{ij} w_{ij} s_i s_j |_{\text{Data}}.$$

Since we are looking for the weights, the updating rule should be equivalent to the gradient

$$\frac{\partial \Delta \log p}{\partial w_{ij}} \sim \langle s_i s_j \rangle_{\text{data}} - \langle s_i s_j \rangle,$$

We could update our weights using a rule compatible with the gradient of the probability.

$$\Delta w_{ij} \sim \langle s_i s_j \rangle_{\text{data}} - \langle s_i s_j \rangle. \quad (6.1)$$

It's easily noticed that the first term is basically Hebbian learning rule, where similar activities enhances the weight. The second term is the some unlearning rule where we have to reduce some weights to relax to the actual working network. Simply put, we kill some connects that have negative effects on our learning network.

Without Hidden Units

In principle, a learning process can be as simple as a one on one map from the data to all the neurons in Boltzmann machine.

Suppose we have a data set of an image which has 10 pixels and each pixel can take values of 0 and 1. We simply construct a network of 10 neurons. To remember the most prominent features of the image with this network, we update the weights and bias of the network to minimize the energy. Once done with minimization, the network we have could be used to generate similar images with similar features.

With Hidden Units

The complexity of the previous model seems to be low. To introduce more degree of freedom, we introduce the hidden units.

Hinton and Sejnowski worked out an algorithm for Boltzmann machine in 1983. The energy based learning rule for Boltzmann machine has two phases, the clamped phase (Phase+) and free phase (Phase-).

1. Clamped phase: we attach the data to the visible units and initialize the hidden units to be some random states.
 1. Then we update the hidden units so that the energy is minimized with clamping.
 2. We calculate the average $s_i s_j$ over all pairs of units, i.e., $\langle s_i s_j \rangle_{\text{data}}$.
 3. Repeat for all data sets of training.
2. Free phase: no clamping and all units are the same and mutable.
 1. Relax the state of the units to low energy state.
 2. Find out $\langle s_i s_j \rangle_{\text{model}}$.
 3. Repeat N times.

With $\langle s_i s_j \rangle_{\text{data}}$ and $\langle s_i s_j \rangle_{\text{model}}$ found, we use the update rule (??).

For some reference, Hinton has a set of lectures on [Coursera](#). He also has a lecture for more effective algorithms: [More efficient ways to get the statistics](#).

6.3.3 Minimizing Energy of Ising Model is Hebbian Learning

Hebbian Learning Rule

Simply put, neurons act similarly at the same time would be more likely to be connected.

A energy minimization procedure would be the same as Hebbian learning rule. Suppose we pick out two spins, $s_3 = 1$ and $s_8 = 1$, the connected weight would be positive in order to have lower energy $-J_{38}s_3s_8 = -J_{38}$. For spins with different signs, negative weight would be the choice to make sure the energy is lower. This is similar to Hebbian learning rule.

Programming

To code a Boltzmann machine, we need a protocol.

COLLECTIVE INTELLIGENCE

7.1 Collective intelligence References

1. **MIT Center for Collective Intelligence.** “Our basic research question is: How can people and computers be connected so that—collectively—they act more intelligently than any person, group, or computer has ever done before?”
2. [Collective intelligence@Wikipedia](#)

SOME TOPICS

Some random but interesting topics.

8.1 Mean-field Method

Mean-field is a concept in statistical mechanics to approximately treat the many body problem. It is an effective theory.

8.1.1 Paramagnetic Material

In paramagnetic materials, the “polarized” spins $\{s_i\}$ determine the overall “magnetization” of the material. When the spins are all aligned, we get the maximum “magnetization”. If the directions of the spins are completely random, the material is not “magnetized”. In other words, we can use a quantity $m = \langle s \rangle$, which is called magnetization, to represent how well the overall “magnetization” of the material is.

For simplicity, we assume that the spins can only be up ($s=1$) or down ($s=-1$).

The idea of mean-field is to treat each spin as a mean-field m and the fluctuations around this mean-field. Mathematically speaking

$$s_i = m + \delta s_i,$$

where $\delta s_i = s_i - m$ is the so called fluctuations. This treatment of the material requires the material to be homogeneous. Under the condition of homogeneous material, we can think of all the spins are the same. Thus we can look at one spin s_α and all other spins become a background environment of this particular spin. This background environment is called mean-field which provides a potential field to interact with.

On the other hand, the average of spins $\langle s \rangle$ is equivalent to the ensemble average of this particular spin $\langle s_\alpha \rangle$

$$\langle s \rangle = \langle s_\alpha \rangle_{ensemble} = \sum P(E_\alpha) s_\alpha,$$

where $P(E_\alpha)$ is calculated by the Boltzmann distribution,

$$P(E_\alpha) = \frac{e^{\beta E_\alpha}}{\sum e^{\beta E_\alpha}}.$$

So we need to calculate the energy of the spin first. Using the Hamiltonian for spins

$$E_\alpha = -s_\alpha H_{total},$$

where H_{total} is the total magnetic field including the mean-field. This energy finally becomes

$$\begin{aligned} E_\alpha &= -s_\alpha \left(H_{ext} + J \sum_{i, near \alpha} s_i \right) \\ &= -s_\alpha \left(H_{ext} + J \sum_{i, near \alpha} (m + (s_i - m)) \right) \\ &= -s_\alpha \left(H_{ext} + Jzm + J \sum_{i, near \alpha} (s_i - m) \right), \end{aligned}$$

where z is the number of spins interacting with s_α .

Mean-field theory assume a homogeneous background, the summation of fluctuations, $\sum_{i, near \alpha} (s_i - m)$ should be 0.

Then energy is calculated as

$$E_\alpha = -s_\alpha (zJm + H).$$

Using this we get the ensemble average of spin α ,

$$\begin{aligned} \langle s_\alpha \rangle &= \sum_{s_\alpha} s_\alpha \frac{e^{\beta E_\alpha}}{\sum e^{\beta E_\alpha}} \\ &= \tanh(\beta(zJm + H)). \end{aligned}$$

By definition, we know that

$$m = \langle s \rangle = \langle s_\alpha \rangle_{ensemble} = \tanh(\beta(zJm + H)).$$

Such a transcendental equation tells us the magnetization of the material.

8.1.2 Neuroscience

Mean-field theory is also applied to networks of neurons. Since it is an effective theory, we can only talk about the statistical features of neuron networks as mean-field is applied.

This method is explained in [Deco2008]. This section is an interpretation of it if not a rewritten.

Two Phase Spaces

Similar to the μ space (6 D) and Γ space (6N D) in statistical mechanics, it takes a lot of dimensions if we would like to describe a neuron network using a point in a phase space (Γ space of network).

To see this, consider a network of N neurons, each neuron requires ζ variables $\{\xi_1, \xi_2, \dots, \xi_\zeta\}$ for a complete description (say, PSP V , current I , \mathcal{T} , as mentioned in [Deco2008]). If we would like to describe the network using these ζ variables, it is represented as a bunch of points in such a phase space, which we name μ space. On the other hand, we can construct a $N\zeta$ dimensional phase space, where each dimension is a variable of a certain neuron. Suppose the neurons are given ids i . The dimensions are $\xi_{j,i}$ which stands for the ξ_j variable of neuron i .

We could imagine how convoluted the new phase space (Γ space) has become.

In any case, the evolution of the network is mapped to the motion of a point in Γ space or motion of N points in μ space.

Equation of Motion

The dynamics of the system is described using Fokker-Planck equation,

$$\partial_t p = -\nabla \cdot (f - D\nabla)p,$$

where p is the density of points in phase space. This is a statistical formalism.

Drift and Diffusion

The two terms on the RHS are dubbed as drift term and diffusion term for the physics they are describing.

[Deco2008] uses leaky integrate-and-fire model, which is used to find out the evolution of the membrane potential V_i of neuron i , which has a solution

$$V_i = V_L + \sum_{j=1}^N J_{ij} e^{-(t-t_j^{(k)})/\tau} \sum_k \Theta(t - t_j^{(k)}),$$

where Θ is the Heaviside function. The complication of this model comes from the fact that all neurons are potentially coupled.

Leaky Integrate-and-fire Model

LIF is a model utilizing capacitors and resistors to describe the evolution of membrane potential. The point of a capacitor is to reset the potential when it reaches the threshold.

The EoM is

$$\tau \frac{d}{dt} V = -(V - V_L) + IR,$$

where I is the total synaptic current input of the neuron, V_L is the leak.

For a neuron i in a network, I comes from all the synaptic delta pulse inputs of other neurons j ,

$$I_i R = \tau \sum_{j=1}^N J_{ij} \sum_k \delta(t - t_j^{(k)}),$$

where $t_j^{(k)}$ denotes the time for k th spike of neuron j .

The equation can be solved using Green's function method.

Following the idea of phase space, the authors of [Deco2008] derived Fokker-Planck equation for spiking neuron models. We could imagine that the Γ space for LIF models would be a N dimensional one $\{V_i\}$, where N is the number of neurons. On the other hand, μ space description requires only 1 D, which is potential. In μ space, we define a probability density $p(v, t)$, which gives us the probability of neuron falls in a potential range $[v, v + dv]$ when multiplied by the size of the phase space dv .

The Chapman-Kolmogorov equation describes the evolution of such a density function

$$p(v, t + dt) = \int_{-\infty}^{\infty} p(v - \epsilon, t) \rho(\epsilon | v - \epsilon) d\epsilon,$$

where $\rho(\epsilon | v - \epsilon)$ is the probability of neuron to be with potential v at $t + dt$ given neuron potential at t to be $v - \epsilon$. But this integral form is not always useful. Kramers-Moyal expansion of it gives us the differential form

$$\partial_t p(v, t) = \sum_{k=1}^{\infty} \frac{(-1)^k}{k!} \frac{\partial^k}{\partial v^k} \left(p(v, t) \lim_{dt \rightarrow 0} \frac{\langle \epsilon^k \rangle_v}{dt} \right), \quad (8.1)$$

where

$$\langle \epsilon^k \rangle_v = \int_{-\infty}^{\infty} \epsilon^k \rho(\epsilon|v) d\epsilon,$$

which is an average.

The plan is to find method to calculate $\langle \epsilon^k \rangle_v$ which is an average with respect to v .

Mean-field Approximation

By assuming each neurons are identical in this network, we face the same situation as in the paramagnetic material. By saying so we are going to assume that the time average (system average, in paramagnetic materials) is the same as the ensemble average, which treats all other neurons as a background and consider only one neuron.

The key of this practice is that $\langle \epsilon^k \rangle_v = \langle \epsilon^k \rangle_{ensemble}$. On the other hand, we know that $\langle \epsilon^k \rangle_{ensemble} = dV$, if we find this dV that calculates the change of potential.

We need to find a quantity that is useful to describe the network, which should be similar to our magnetization m . The choice can either be average potential of each neuron V_{avg} or the total membrane potential $V = NV_{avg}$. The authors in [Deco2008] choose the average membrane potential.

We denote mean firing rate as A . The equation of V is

$$dV = \langle J \rangle N A dt - \frac{V - V_L}{\tau} dt,$$

where $\langle J \rangle$ is the average synaptic weight and mean firing rate or population activity is

$$A = \lim_{dt \rightarrow 0} \frac{\text{number of spikes within time } [t, t + dt]}{dt} \frac{1}{N}.$$

Since $\epsilon = dV$, we plug it in to Eq. (??). To save keystrokes, we calculate the moments first.

$$\begin{aligned} M^{(k)} &= \lim_{dt \rightarrow 0} \frac{1}{dt} \langle \epsilon \rangle_v \\ &= \lim_{dt \rightarrow 0} \frac{1}{dt} \langle \epsilon \rangle_{ensemble} \\ &= \lim_{dt \rightarrow 0} \frac{1}{dt} dV \\ &= \dots \\ &= \langle J^k \rangle N A. \end{aligned}$$

In the derivation we dropped all terms that contains order of infinitesimal variables $\mathcal{O}(dt^2)$.

The diffusion approximation will help us drop the $k > 2$ moments. This is a Taylor expansion anyway.

A Simple Argument for Diffusion Approximation?

There should be one.

Regardless of real understanding or not,

$$\begin{aligned} M^{(1)} &= \frac{\mu(t)}{\tau} - \frac{v - V_L}{\tau} \\ M^{(2)} &= \frac{\sigma^2(t)}{\tau}, \end{aligned}$$

are related to the drift coefficient $\mu(t)$ and the diffusion coefficient $\sigma(t)$.

Then we obtain the Fokker-Planck equation

$$\partial_t p(v, t) = \frac{1}{2\tau} \sigma^2(t) \partial_v^2 p(v, t) + \partial_v \left[\frac{v - V_L - \mu(t)}{\tau} p(v, t) \right],$$

where $\sigma(t)$ is diffusion coefficient and $\mu(t)$ is drift coefficient.

8.1.3 References and Notes

8.2 Spin Glass

In the realm of physics, Ising model has been and probably will be an useful model for our intuition. Naively speaking, our world is full of local interactions yet the consequences are very likely to be long range through long range correlations. The famous Ising model is a key model of such concepts.

Ising model deals with a system of spins and their interactions. To build up the dynamics of a system, we need the configuration of it as well as the interactions. In Ising mode, a bunch of spins are described by $\{s_i\}$, where s_i could be scalars that take discrete values or vectors phenomenologically. For simplicity we take the path of discretized values, $s_i = 1, -1$. The interactions are two body interactions is related to the description of energy

$$H = - \sum_{i,j} J_{i,j} s_i s_j,$$

where $J_{i,j}$ indicates the strength of the interactions between the two spins.

8.3 Binary Human Model

Han has been discussing with me about building human dynamics by making analogy to neural systems.

The idea is to treat every human as an activation function which is connected to each other to form a large network. The trouble was how to deal with information input and output from outside of the group which is similar to external stimulation in neural systems.

8.3.1 A Simplified Model

Here we develop a model for wechat groups instead of general human dynamics. We assume everyone share all the information in the group.

There are several aspects to be defined.

0. One way to simplify the model is to set up a external pool of information and opinions which we denote as \mathcal{P} .
1. An opinion is plugged into the pool \mathcal{P} initially.
2. Everyone will take information and output information to this pool \mathcal{P} .
3. Each human would either output null (being silent) or 1 (agreement) or 0 (disagreement). Each of the outputs in the pool is denoted as P_i .
4. The output of each human depends on the pool, the nature of the human, and the relation with other people. We approximate each human using a binary input/output function.

Such a general model seems to be good enough. However, we have to define all different types of human first.

1. Being average: takes in all the opinions of other participants and work out an average.

$$\Theta\left(\frac{\sum_1^N P_i}{N} - 0.5\right),$$

where Θ is the step function or Heaviside function.

2. Being extreme: exaggerates whatever is obtained from average

$$\Theta\left(\alpha \frac{\sum_1^N P_i}{N} - 0.5\right),$$

where α is the extreme factor. For linear summations (euclidean metric), this is effectively pessimism ($\alpha < 1$) or optimism ($\alpha > 1$).

3. Synchronized to other people. The simplest case is to synchronize to one of the members in the group.

Numerical Model

We need to build a program to calculate an example and see if it makes sense.

```
import numpy as np
```

```
# Define Heaviside function; returns 0 if x<0.5, otherwise returns 1

def hsh(x):
    #return (0.5 * (np.sign(x-0.5) + 1))
    return int(np.piecewise(x, [x < 0.5, x >= 0.5], [0, 1]) )
```

```
# Test of functions

print hsh(0.1), hsh(0.5), hsh(0.6)
```

```
0 1 1
```

In this example, we simulate the opinion stream of each person as well as the whole system which is determined by the pool.

For convinience we establish a group of 4 person.

```
# Define the pool

pool = np.random.randint(2, size=4).tolist();
print pool
```

```
[1, 0, 0, 1]
```

```
# Define four characters of the group

## First character is an averagist

def avgist(x):

    return hsh( np.sum(x)/len(x) )

## The second character is pessimist
```

```
def pesist(x,alpha=0.6):

    return hsh( alpha * np.sum(x)/len(x) )

## The third character is a optimist
def optist(x,alpha=2):

    return hsh( alpha * np.sum(x)/len(x) )

## The fourth synchronizes with the first character, i.e., averagist

def synist(x):

    return x[0]
```

```
# Test characters

print avgist(pool), pesist(pool), optist(pool), synist(pool)
print np.array([avgist(pool), pesist(pool), optist(pool), synist(pool) ])
```

```
0 0 1 1
[0 0 1 1]
```

```
# Define iterator

def iter(initpool,n): # initpool: the initial pool; n: number of iterations

    poolM = np.array(initpool);
    stateM = np.array([poolM]);

    for i in range(n):
        poolM = np.array( [ avgist(poolM), pesist(poolM), optist(poolM),
↪synist(poolM) ] )
        stateM = np.append( stateM, np.array([poolM]) ,axis=0 )

    return stateM
```

```
print(iter([1,1,0,1],10), \
      iter([1,1,0,1],10), \
      iter([1,0,1,1],10)
      )
```

```
(array([[1, 1, 0, 1],
       [0, 0, 1, 1],
       [0, 0, 1, 0],
       [0, 0, 0, 0],
       [0, 0, 0, 0],
       [0, 0, 0, 0],
       [0, 0, 0, 0],
       [0, 0, 0, 0],
       [0, 0, 0, 0],
       [0, 0, 0, 0],
       [0, 0, 0, 0]]), array([[1, 1, 0, 1],
       [0, 0, 1, 1],
```

```

[0, 0, 1, 0],
[0, 0, 0, 0],
[0, 0, 0, 0],
[0, 0, 0, 0],
[0, 0, 0, 0],
[0, 0, 0, 0],
[0, 0, 0, 0],
[0, 0, 0, 0],
[0, 0, 0, 0],
[0, 0, 0, 0]], array([[1, 0, 1, 1],
[0, 0, 1, 1],
[0, 0, 1, 0],
[0, 0, 0, 0],
[0, 0, 0, 0],
[0, 0, 0, 0],
[0, 0, 0, 0],
[0, 0, 0, 0],
[0, 0, 0, 0],
[0, 0, 0, 0],
[0, 0, 0, 0],
[0, 0, 0, 0],
[0, 0, 0, 0],
[0, 0, 0, 0],
[0, 0, 0, 0],
[0, 0, 0, 0]]))

```

8.4 Factor Graph

In the propagation of information, we are usually dealing with such relations that some system (factor) f is related to some inputs (variable) x .

Factor graph builds such networks of relations. In such a graph, we use boxes to represent factors, and circles to represent variables.

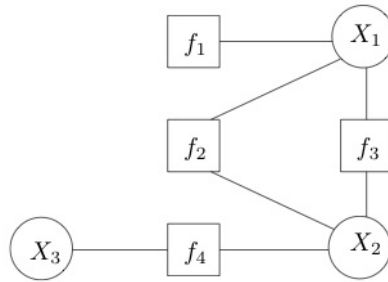


Fig. 8.1: From Factor Graph @ Wikipedia.

Suppose we need to find out the probability of factor f_4 , we have to take in the two variables X_2 and X_3 .

One important question is that we can actually calculate the probability of a specific configuration of the variables, which is

$$P \propto f_1(X_1) + f_2(X_1, X_2) + f_3(X_1, X_2) + f_4(X_2, X_3).$$

In general, we can assign weights w_i to each factors f_i , so that the probability of each set of variables is

$$P(X_1, X_2, X_3, X_4) \propto w_1 f_1(X_1) + w_2 f_2(X_1, X_2) + w_3 f_3(X_1, X_2) + w_4 f_4(X_2, X_3).$$

The normalization is the partition function, i.e., the summation of all the possible combinations of variables,

$$Z = \sum_{\{X_1, X_2, X_3, X_4\}} P(X_1, X_2, X_3, X_4).$$

As an example, we consider X_i is 0 or 1. The combinations are (partially) listed below.

Table 8.1: List of Possible Configurations of Variables

X_1	0	1	1	1	...
X_2	0	0	1	1	...
X_3	0	0	0	1	...
X_4	0	0	0	0	...

8.4.1 References and Notes

1. Factor Graph @ Wikipedia.
2. DeepDive Docs.

SOME RANDOM THOUGHTS

9.1 Pan-Intelligence

Biological neural networks give us human intelligence. Are biological neural networks special out of the numerous different kinds of network systems? I prefer to believe they are not special.

So what is a list of all systems that could function as a neural network?

1. Social networks. We could model the society as an information processing network. It is extremely complicated, at least more complicated than the models in physics.
2. The whole ecological system on the Earth is like a web of lives.
3. Networks that connects time instead of space is also an interesting idea. If we have nodes N_i located at different time t_i and the nodes are connected. Now we have a network which can be analysis. However, for causality reasons we should limit the network to be directed.

9.2 Why Can We Understand the Universe?

Our brain contains significantly less degrees of freedom than the whole universe. Yet we are developing very good understanding of whole universe.

One of the interesting points is that the universe doesn't go through all the possible states of its configuration. There are many constraints on the universe that forbids it from developing into many other states. Thus we could understand the restrictions thus grasp the key of the state of the universe.

In some sense, intelligence should have this ability to find the restrictions of the environment and develop understanding of how the environment changes. Without the understanding of the restrictions of the enormously complex environment, it won't be possible to predict.

REFERENCES

ACKNOWLEDGEMENT

I am grateful to the authors of Sphinx.

My wife Han Lu helps me a lot in this set of learning notes. She provided topics, ideas, and organized a reading club on the topic of neuroscience, all of which set my path to learn more about intelligence.

BIBLIOGRAPHY

- [Tetzlaff2012] Tetzlaff, C., Kolodziejski, C., Markelic, I., & Wörgötter, F. (2012). Time scales of memory, learning, and plasticity. *Biological Cybernetics*, 106(11–12), 715–726. <https://doi.org/10.1007/s00422-012-0529-z>
- [Tetzlaff2013] Tetzlaff, C., Kolodziejski, C., Timme, M., Tsodyks, M., & Wörgötter, F. (2013). Synaptic Scaling Enables Dynamically Distinct Short- and Long-Term Memory Formation. *PLoS Computational Biology*, 9(10), e1003307. <https://doi.org/10.1371/journal.pcbi.1003307>
- [LaFerla2008] Intracellular amyloid- β in Alzheimer’s disease (Figure 3 is a nice illustration of the process), Poster: Amyloid- β and tau in Alzheimer’s disease
- [Jaunmuktane2015] Evidence for human transmission of amyloid- β pathology and cerebral amyloid angiopathy
- [Palop2016] Network abnormalities and interneuron dysfunction in Alzheimer disease
- [Newman2016] Newman, M. E. J. (2016). Community detection in networks: Modularity optimization and maximum likelihood are equivalent. *ArXiv* 1606.02319, 1–8.
- [Zhang2014] Zhang, P., & Moore, C. (2014). Scalable detection of statistically significant communities and hierarchies using message passing for modularity, 111(51), 18144–18149.
- [Boguna2010] Boguna, M., Papadopoulos, F., & Krioukov, D. (2010). Sustaining the Internet with Hyperbolic Mapping. *Nature Communications*, 1(6), 1–8.
- [BrainConnectivity] Brain connectivity
- [Sivakumar2016] Sivakumar SS, Namath AG and Galán RF (2016) Spherical Harmonics Reveal Standing EEG Waves and Long-Range Neural Synchronization during Non-REM Sleep . *Front. Comput. Neurosci.* 10:59. doi: 10.3389/fncom.2016.00059
- [Germentrout] *Mathematical Foundations of Neuroscience* Chapter 6.
- [Lagzi2015] Dynamics of competition between subnetworks of spiking neuronal networks in the balanced state
- [Deco2010] The Dynamical Balance of the Brain at Rest
- [Deco2014] Great Expectations: Using Whole-Brain Computational Connectomics for Understanding Neuropsychiatric Disorders
- [Gerstner2002] Wulfram Gerstner and Werner M. Kistler, *Spiking Neuron Models*, (2002).
- [Deco2008] Deco, G., Jirsa, V. K., Robinson, P. A., Breakspear, M., & Friston, K. (2008). *The Dynamic Brain: From Spiking Neurons to Neural Masses and Cortical Fields*. *PLoS Computational Biology*, 4(8), e1000092.