

Advanced mathematical methods for biology

Jan-Hendrik Schleimer (AG Theoretische Neurophysiologie, Prof. Schreiber)

Contents

1 Preface	2
1.1 (Dis)claimer	3
2 The Floquet theory of the aktion potential	3
2.1 Periodic event	3
2.2 Tonic spikes are limit cycles	3
2.3 Limit cycle stability	4
2.3.1 Eigensystem of the Floquet matrix	6
2.4 Neutral dimension and phase shifts	7
3 Fourier theory	7
3.1 The Fourier base	7
3.2 Existence of the Fourier integral	7
3.2.1 Orthonormality	8
3.2.2 Convolution	8
3.2.3 Derivative	9
4 The continuum limit of a membrane patch	9
4.1 The ion channel as a Markov model	9
4.1.1 The n -state channel	11
4.1.2 Simulation a Jump process	12
4.2 Statistically equivalent diffusion process (Orenstein-Uhlenbeck Process)	13
5 Information theory for the living	14
5.1 The communication process	14
5.2 Source coding, data compression and efficient representation	15
5.3 Channel coding	18
5.4 Information transmission (continuous case)	20
5.5 Linear stimulus reconstruction and a lower bound on the information rate (decoding view)	24
6 Linear response filter	27
6.1 Instantaneous firing rate in the continuum limit	28

6.2	Phase flux = firing rate	29
7	Numerical continuation of fixpoints and orbits	32
7.1	Continuation of fixed points	32
7.2	Local bifurcations: What happens if $\nabla \vec{f}$ is singular?	34
7.2.1	Folds and one-dimensional nullspaces	34
7.3	Stability exchange	36
7.3.1	Extended system	36
7.4	Continuation of boundary value problems and periodic orbits	36
8	PRC near the centre manifold	36
8.1	Dynamics in the centre manifold of a saddle-node	36
8.2	PRC near the centre manifold	38
9	Phase suscetability to channel noise near saddle-node loop bifurcations	39
10	Event time series and phase descriptions	41
10.1	Synchronisation and phase locking	41
10.2	Spike reliability and stimulus entrainment	43
10.3	Inter-spike statistics	45
10.3.1	First passage time (no input and constant noise)	45
10.3.2	Phase dependent noise	45
10.3.3	Moments of the ISI distribution	48
10.4	Moments of the inter-spike interval distribution	48
10.5	Renewal equation	49
10.6	Spike auto-spectrum	51
10.7	Mutual entrainment	51
10.7.1	Spike metric	52
10.7.2	Time to Spike	52
11	Appendix	52
11.1	Novikov-Furutsu-Donsker formula	52
	Bibliography	53

1 Preface

Location:	House 6, Lecture Hall - Campus Nord
Time:	Tue., 16:00-18:00
Contact:	Email
Script:	under construction

It is still a great challenge to come up with quantitative assertions about complex biological systems. Especially, if one aims towards a functional understanding at the cell, tissue or organismic level, it typically leads to quite involved models. Analytical progress can be made with simplifying assumptions and the restriction to limiting cases.

This course will discuss a potpourri of mathematical methods ranging from analytical techniques to numerical methods and inform the participant on how to apply them to answer biological questions and have enormous fun in doing so.

The mathematical techniques encompass stochastic systems, numerical bifurcation analysis, information theory, perturbation theory and Fourier analysis. The biological examples are chosen mostly from neurobiology, sensory ecology, but also bacterial communication and evolution.

1.1 (Dis)claimer

This scriptum shall evolve. It will have flaws (didactic ones and outright errors) but you, the reader, student, search bot, provide the selective pressure to improve it. Please write to the lecturer if you find faults of any kind with it.

2 The Floquet theory of the action potential

2.1 Periodic event

Of many natural systems we observe only some drastic events.

The clock regular pulses of pulsars [4]

Action potentials There is a voltage threshold. The dynamics is described by *conductance based* equations.

Wake up time Melatonin level below threshold (suprachiasmatic nucleus)

cell division Gap phase 1 → synthesis → gap phase 2 → mitosis

2.2 Tonic spikes are limit cycles

Many interesting differential equations describing biological dynamics¹ can be cast into the following form

$$(1) \quad \dot{\vec{x}} = \vec{F}(\vec{x}, \vec{\theta}) + \vec{\eta}(\vec{x}, t)$$

where \vec{x} is a vector of state variables, *e.g.*, voltage and kinetic variables. \vec{F} is the time independent (homogeneous) flow field. $\vec{\eta}$ is a (often assumed small),

¹Dynamics just means there are derivatives in our equations, not static relations

possibly time dependent (inhomogeneous) perturbation to the system. $\vec{\theta}$ are system's parameter. The change of solutions in response to variations in system's parameter will be studied later.

Example: Hodgkin & Huxley equations From Kirchhof's law of current conservation at the membrane follows a dynamics for the membrane voltage $c\dot{v} + \sum_k^K I_k^{\text{ion}}(v, \vec{m}_k) = I_{\text{in}}$. The first term is the capacitive (pseudo current), followed by a sum of currents from voltage dependent ion channels and the applied input current I_{in} . The ion channel dynamics is governed by kinetic equations, *i.e.*, chemical reactions between *open* and *closed* states. The reactions often follows first order kinetic equations $\tau_k(v)\dot{m}_k = \underline{M}_k^{(\infty)}(v) - \vec{m}_k$. $\tau_k(v)$ is the matrix of kinetic time constants and $\underline{M}_k^{(\infty)}(v)$ are the steady state activation curves. τ and $\underline{M}^{(\infty)}(v)$ are typically diagonal if the channels are independent. These equations will be derived more formally in Sec. XX. The state vector is then $\vec{x} = [v, m_{11}, \dots, m_{K1}, \dots]^\dagger$. In the absence of perturbations Eq. (1) becomes a homogeneous equation

$$(2) \quad \dot{\vec{x}} = \vec{F}(\vec{x}, \vec{\theta})$$

and we assume the existence of a P -periodic limit cycle solution, $\vec{x}_{\text{LC}}(t) = \vec{x}_{\text{LC}}(t + P)$, also known as periodic orbit.

A neurobiological dogma states that action potentials follow the *all-or-nothing* principle². This can be construed to mean that the exact shape of the action potential does not matter³ and that information is stored in the exact spike times⁴. It also means that the action potential should be stable under perturbations, but that inputs ought to be able to shift the occurrence of spikes in time. If you want, from the *dogma* and the intend to code into spike times follow two *desiderata*:

- I) The action potential must be *stable* in amplitude,
- II) yet *neutrally stable* in phase.

To wit, II. just means that stimulus induced phase shifts should neither decay nor blow up.

2.3 Limit cycle stability

Stability of is probed by studying small perturbation to an invariant set solution. Our invariant set is the limit cycle (periodic orbit) \vec{x}_{LC} . Assuming there was a *small* perturbation to the system the solution can be decomposed as

²Fortunately the jelly fish *Aglanta digitalae* does not care much for dogmas and encodes its swimming patterns in different action potential shapes

³Actually it does, if you bring energetic considerations into baring.

⁴There is also plenty of coding in graded potentials going on for example in some of *Drosophila melanogaster's* neuron or *your* retina. *C. elegans* seems to do completely without spikes.

$$(3) \quad \vec{x}(t) = \vec{x}_{\text{LC}}(t) + \vec{y}(t)$$

with $\forall t : \|\vec{y}(t)\| < \epsilon$ some “small” perturbation to the orbit. What small, *i.e.*, ϵ is we do not want to say now, maybe later, lets see . . .

Assuming the perturbation was transient (only in initial conditions) and the system is homogeneous again we plug the Ansatz of Eq. (3) into Eq. (2) and get

$$\frac{d}{dt}\vec{x}_{\text{LC}}(t) + \dot{\vec{y}}(t) = \vec{F}(\vec{x}_{\text{LC}}) + \underbrace{\nabla\vec{F}(\vec{x}_{\text{LC}})}_{\underline{J}(\vec{x}_{\text{LC}}(t))} \cdot \vec{y}(t)$$

The *Jacobi matrix* evaluated on the limit cycle can be written as a function of time $\underline{J}(t) = \nabla\vec{F}(\vec{x}_{\text{LC}}(t))$. Note that since the limit cycle solution is P -periodic, so is $\underline{J}(t) = \underline{J}(t + P)$.

Identifying the limit cycle solution above we are left with the *first variational equation* of Eq. (2)

$$(4) \quad \dot{\vec{y}}(t) = \underline{J}(t)\vec{y}(t).$$

Hence, one needs to study of linear system with periodic coefficients. One solution of Eq. (4) can be guessed, let us try the time derivative of the orbit, $\frac{d\vec{x}_{\text{LC}}}{dt}$

$$(5) \quad \frac{d}{dt}\left(\frac{d\vec{x}_{\text{LC}}}{dt}\right) = \frac{d}{dt}\vec{F}(\vec{x}_{\text{LC}}) = \nabla\vec{F}(\vec{x}_{\text{LC}})\frac{d}{dt}\vec{x}_{\text{LC}} = \underline{J}(t)\frac{d\vec{x}_{\text{LC}}}{dt}.$$

So it is a solution alright, and it happens to be a P -periodic solution. This solution is called the *Goldstone mode*. But for arbitray intial conditions not all solutions should be periodic.

Def: Floquet Ansatz According to Floquet theory the solution to Eq. (4) can be written in the form of a P -periodic similarity matrix⁵ and an matrix exponential

$$(6) \quad y(t) = \underline{R}(t)e^{t\underline{A}}.$$

For the proof please consult (Chicone 2006), here we are just goint to work with this as an Ansatz. The constant matrix \underline{A} is called the *Floquet matrix*.

Recall the matrix exponential

Def: Matrix Exp Let $\underline{A} \in \mathcal{O}^{n \times n}$ then $\exp \underline{A} = \sum_{k=0}^{\infty} \frac{1}{k!} \underline{A}^k$ A useful corollary of this definition is that the eigen vectors of an exponentiated matrix are the same as those of the original and eigenvalues become exponentiated. If λ_i , \vec{w}_i are the eigenvalue, eigenvector pairs of the matrix \underline{A} , *i.e.*, $\underline{A}\vec{w}_i = \lambda_i\vec{w}_i$ then by using this identity k -times

$$(7) \quad e^{\underline{A}}\vec{w}_i = \left(\sum_{k=0}^{\infty} \frac{1}{k!} \underline{A}^k\right) \vec{w}_i = \sum_{k=0}^{\infty} \frac{1}{k!} \lambda_i^k \vec{w}_i = e^{\lambda_i} \vec{w}_i$$

Inserting the Floquet Ansatz into Eq. (4) yields

⁵For now this means that its an invertable matrix; $\forall t : \exists \underline{R}^{-1}(t)$

$$\dot{\underline{R}}e^{t\Lambda} + \underline{R}(t)\underline{\Lambda}e^{t\Lambda} = \underline{J}(t)\underline{R}(t)e^{t\Lambda}.$$

Which by multiplying with $e^{-t\Lambda}$ results in a dynamics equation for similarity matrix \underline{R} .

$$(8) \quad \dot{\underline{R}} = \underline{J}(t)\underline{R}(t) - \underline{R}(t)\underline{\Lambda}.$$

Remember that \underline{R} was invertable so one can also derive an equation for the inverse

$$(9) \quad \frac{d}{dt}\underline{R}^{-1} = \underline{R}^{-1}\dot{\underline{R}}\underline{R}^{-1} = \underline{\Lambda}\underline{R}^{-1}(t) - \underline{R}^{-1}(t)\underline{J}(t).$$

2.3.1 Eigensystem of the Floquet matrix

The Floquet matrix, $\underline{\Lambda}$, is constant, though not necessarily symmetric. Hence it has an orthonormal left and right eigensystem

$$(10) \quad \underline{\Lambda}\vec{w}_i = \mu_i\vec{w}_i \text{ and } \vec{z}_i\underline{\Lambda} = \mu_i\vec{z}_i \text{ with } \vec{z}_i \cdot \vec{w}_j = \delta_{ij}$$

One can define a “rotated” eigensystem

Def: (“Rotated” eigensystem): $\vec{W}_i(t) = \underline{R}(t)\vec{w}_i$ and $\vec{Z}_i(t) = \vec{z}_i\underline{R}^{-1}(t)$. For which *ipso facto* obeys

$$(11) \quad \forall t : \vec{Z}_i(t) \cdot \vec{W}_j(t) = \delta_{ij}.$$

$\vec{W}_i(t)$ and $\vec{Z}_i(t)$ are also called the *Floquet modes*.

If we project the eigenvectors on the Eqs. (8) and (9) and use this definitions we get

$$(12) \quad \frac{d}{dt}\vec{W}_k = (\underline{J}(t) - \mu_k\underline{I})\vec{W}_k(t)$$

and

$$(13) \quad \frac{d}{dt}\vec{Z}_k = (\mu_k\underline{I} - \underline{J}^\dagger(t))\vec{Z}_k(t).$$

If one projects the general solution $y(t)$ from Eq. (6) on the the adjoint Floquet modes

$$\vec{Z}_k(t) \cdot \vec{y}(t) = \vec{z}_k\underline{R}^{-1}(t)\underline{R}(t)e^{t\Lambda} = \vec{z}_ke^{t\Lambda} = \vec{z}_ke^{t\mu_k}$$

If $\nu_k < 0$ the perturbation decays exponentially in this rotated coordinate frame.

Poincare Section bla

Note that if $\nu_0 = 0$, then according to Eq. (5)

$$(14) \quad \vec{W}_0(t) = \frac{d}{dt}\vec{x}_{\text{LC}}(t)$$

is the Goldstone mode, and from Eq. (11)

$$(15) \quad \vec{Z}_0(\phi) \cdot \frac{d}{dt}\vec{x}_{\text{LC}}(\phi) = 1$$

2.4 Neutral dimension and phase shifts

The evolution of the phase of Eq. (1) is given by

$$\frac{d\phi}{dt} = \nabla\phi(\vec{x}) \cdot \frac{d\vec{x}}{dt}.$$

To first order this is

$$\frac{d\phi}{dt} = \nabla\phi(\vec{x}_{\text{LC}}) \cdot \frac{d\vec{x}_{\text{LC}}}{dt} = \nabla\phi \cdot \vec{F}(\vec{x}_{\text{LC}}) + \nabla\phi \cdot \vec{\eta}(\vec{x}_{\text{LC}}, t).$$

There are several ways to define a phase (Hilbert transform, linear interpolation, ...). A *desiderata* could be to have a linear phase increase in the unperturbed case ($\vec{\eta} = 0$), say $\phi(t) = f_0 t$. [... proto-phase]. From this *desiderata* it follows that one must have $\forall t : \nabla\phi \cdot \vec{F}(\vec{x}_{\text{LC}}) = f_0$. Given Eq. (16), this is easily achieved with the following identification

$$(16) \quad \nabla\phi = \vec{Z}_0(\phi) f_0 = \vec{Z}(\phi)$$

The *input-output (I/O) equivalent phase oscillator* to Eq. (1) can then be written as

$$(17) \quad \dot{\phi} = f_0 + \vec{Z}(\phi) \cdot \vec{\eta}(t)$$

A spike-train can be written as

$$(18) \quad y(t) = \sum_k \delta(t - t^{\text{SP}}) = \sum_k \delta(\phi(t) - k)$$

3 Fourier theory

To proceed with the analysis we need some results from Fourier theory.

3.1 The Fourier base

One may think of the Fourier transform as a projection onto a new basis $e_\omega(t) = e^{i\omega t}$. Define the projection of a function as

$$F(\omega) = \langle e_\omega \cdot f \rangle = \int_{-\infty}^{\infty} dt e_\omega^*(t) f(t)$$

The inverse Fourier transform is a projection onto $e_t^*(\omega)$

$$f(t) = \langle e_t^* \cdot F \rangle = \int_{-\infty}^{\infty} d\omega e_t(\omega) F(\omega)$$

3.2 Existence of the Fourier integral

Usually all transcendental pondering about the existance of mathematical objects is the subject of pure math and should not bother us too much (we trust the work as been done properly). But in the Fourier transform case it motivates a subject we need to address mainly because of δ -functions, which are heavily used

in theoretical neurobiology, because they are reminiscent of action potentials or indicate the time of such a spike event. What does it mean for the Fourier transform to exist? It means that the integrals involved in the definition of the Fourier transform converge to finite values. OK. So let us look at the magnitude of the transform of a function f . It can be upperbound by the Cauchy-Schwarz inequality

$$|\langle e_w \cdot f \rangle| = \left| \int_{-\infty}^{\infty} dt e^{-i\omega t} f(t) \right| \leq \int_{-\infty}^{\infty} \underbrace{|e^{-i\omega t}|}_{=1} |f(t)| dt = \int_{-\infty}^{\infty} |f(t)| dt.$$

This means that if one assumes the function $f(t)$ is *absolute integrable*

$$\int_{-\infty}^{\infty} |f(t)| dt < \infty,$$

then the Fourier integral exists – hurray. Of course, the same works for the integral involved in inverse Fourier transform. Note that this is an implication in one direction. All functions satisfying absolute integrability are lumped together in $L^1(\mathbb{R})$.

The bad news is that applying a Fourier transform to one of the members of $L^1(\mathbb{R})$ can throw you out of it. And then? Well, luckily there is the set of *Schwartz functions*, which is closed under the Fourier transform.

3.2.1 Orthonormality

The issue that absolute integrability is insufficient already manifests when trying to Fourier transform its own basis, since $\int_{-\infty}^{\infty} dt |e^{i\omega t}| = \infty$. Let's give it a name anyway

$$\langle e_\omega \cdot e_\nu \rangle = \int_{-\infty}^{\infty} dt e^{i(\nu - \omega)t} = \delta(\nu - \omega)$$

It follows that the delta function is symmetric

$$\delta(\omega) = \int_{-\infty}^{\infty} dt e^{i\omega t} = \int_{-\infty}^{\infty} dt e^{-i\omega t} = \delta(-\omega)$$

Note also that

$$(19) \quad \langle e_\omega \cdot 1 \rangle = \delta(\omega)$$

3.2.2 Convolution

The convolution of two functions is defined as

$$h(t) = (f * g)(t) = \int_{-\infty}^{\infty} dr f(t - r)g(r)$$

$$\begin{aligned} H(\omega) &= \int dt e^{i\omega t} \int dr f(t - r)g(r) = \int dr g(r) \int dt e^{-i\omega(t+r)} f(r) \\ &= \int dr g(r) e^{-i\omega r} \int dt e^{-i\omega t} f(t) = G(\omega)F(\omega) \end{aligned}$$

3.2.3 Derivative

What is $\delta'(t)$?

$$(\delta' * f(t)) = \int_{-\infty}^{\infty} \delta'(r)f(t-r)dr = [\delta(r)f(t-r)]_{r=-\infty}^{\infty} - \int_{-\infty}^{\infty} \delta(r)f'(t-r)dr = -f'(t)$$

Hence

$$(20) \quad (\delta' * f(t)) = -f'(t)$$

4 The continuum limit of a membrane patch

Motivation and aim

In this lecture ion channels are introduced as stochastic devices floating in the membrane of a nerve cell. It should motivate why the analysis techniques and models introduced in this lecture need to deal with fluctuations and noise. (Nerve) cells produce stochastic processes on several levels:

1. The ion channels in their membrane stochastically jump between conformations, governed by *Master equations*.
2. On a more macroscopic level their membrane voltage fluctuations show properties of coloured noise, well described by diffusion processes and *stochastic differential equations*.
3. The trains of action potentials they emit from *point processes*.

In the stationary state these processes can be subjected to *spectral analysis*.

Note that in 1952, the first equations describing membrane voltage dynamics where the deterministic rate equations by Hodgkin & Huxley (Hodgkin and Huxley 1952). Only in 1994 Fox and Lu derived these equations from the continuum limit of an ensemble of stochastic ion channels (Fox and Lu 1994). Essentially by doing a *diffusion approximation*.

Since then the nuances of applying diffusion approximations to neurons have been investigated (Linaro, Storace, and Giugliano 2011, Orio and Soudry (2012)) and reviewed (Goldwyn et al. 2011, Goldwyn and Shea-Brown (2011), Pezo, Soudry, and Orio (2014)).

4.1 The ion channel as a Markov model

Proteins change conformation on various triggering signals:

- a change in pH
- a change in the surrounding electric field
- mechanical pressure

Such dynamics can be described by a finite state Markov model which is mathematically described as a Master equation.

Starting with a simple ion channel that has an open conformation, O , in which ions can pass (say K^+ ions) and a closed states, C , which blocks ion flow



We can define a vector of the probability of being open and closed, $\vec{p} = [p_O, p_C]^\dagger$, respectively.

An ionic current produced in the open state would be

$$I^{\text{ion}} = \gamma_{K^+} N_O (E_{K^+} - v)$$

γ_{K^+} and E_{K^+} are the unitary conductance and the Nernst potential respectively. The average such current would be

$$\langle I^{\text{ion}} \rangle = \gamma_{K^+} N p_O (E_{K^+} - v)$$

where N is the total number of channels in the membrane patch under consideration. But what about a particular stochastic realisation of the current, what about the fluctuations around the mean?

If we have N channels than the number of k of them being open is binomially distributed

$$P_O(k) = \binom{N}{k} p_O^k p_C^{N-k}$$

In actuality the channels are part of the membrane dynamical system, where α and β depend at least on v and hence are not constant during a spike. We need an update rule how to get from the probability of being open at time t to the probability of begin open at time $t + dt$. This is given by

$$\begin{pmatrix} p_O(t + dt) \\ p_C(t + dt) \end{pmatrix} = \begin{pmatrix} 1 - \alpha dt & \beta dt \\ \alpha dt & 1 - \beta dt \end{pmatrix} \begin{pmatrix} p_O(t) \\ p_C(t) \end{pmatrix}$$

or in vector form

$$\vec{p}(t + dt) = (\underline{I} + \underline{Q}dt) \vec{p}(t)$$

The infinitesimal limit is

$$\frac{d}{dt} \vec{p} = \underline{Q} \vec{p}$$

With $p_C = 1 - p_O$ we can express one row of this equation as

$$\dot{p}_O = \alpha p_O - \beta(1 - p_O)$$

or

$$\tau \dot{p}_O = p_O^{(\infty)} - p_O \text{ with } \tau = \frac{1}{\alpha + \beta} \text{ and } p_O^{(\infty)} = \frac{\alpha}{\alpha + \beta}$$

which has solution

$$p_O(t) = p_O^{(\infty)} (1 - e^{-t/\tau})$$

Fourier transformation leads to

$$\tau i \omega \tilde{p}(\omega) = p_O^{(\infty)} - \tilde{p}(\omega)$$

or

$$\tilde{p}(\omega) = \frac{p^{(\infty)}}{1+i\tau\omega}$$

$$|\tilde{p}(\omega)|^2 = \tilde{p}(\omega)\tilde{p}^*(\omega) = \frac{p^{(\infty)}}{1+(\tau\omega)^2}$$

a Lorentzian spectrum. Inverse Fourier transform yields the covariance function

$$c(\tau) = \langle p_O(t)p_O(t+\tau) \rangle$$

4.1.1 The n -state channel

Let us try to calculate the statistics of the current originating from an n -state channel (just like in the two state case). *Why* would one do this? The idea, later on is to be able to find a continuous stochastic process that we can simulate and analyse easily.

Let $K(t) \in [1, \dots, n]$ be the realisation of the n -state Markov channel. For example a K^+ -channel with four subunits

$$1 \xrightleftharpoons[\beta]{} 2 \xrightleftharpoons[2\beta]{} 3 \xrightleftharpoons[3\beta]{} 4 \xrightleftharpoons[4\beta]{} 5$$

Assuming the ion channel has n conformations, of which one is conducting, let us further define

$$G = \delta_{1K(t)} = \begin{cases} 1 & : K(t) = 1 \\ 0 & : K(t) > 1. \end{cases}$$

The single channel current at a voltage clamp $v(t) = v$ is then

$$(21) \quad I(t) = \gamma G(t)(E - v).$$

How does it evolve? Define $p_i(t) = P(K(t) = i)$ and $\vec{p}(t) = [p_1, \dots, p_n]^\dagger$, then

$$(22) \quad \frac{d}{dt}\vec{p} = \underline{Q}\vec{p}$$

With formal solution

$$\vec{p}(t) = e^{\underline{Q}t}\vec{p}(0) = \underbrace{(e^{\underline{Q}t})}_{\underline{M}}\vec{p}(0) = \underline{M}^t\vec{p}(0)$$

Use the singular value decomposition, $\underline{M} = \underline{U}\underline{\Sigma}\underline{V}^\dagger$, the matrix power can be written as $\underline{M}^t = \underline{U}\underline{\Sigma}^t\underline{V}^\dagger$. Or (recall Eq. (7))

$$(23) \quad \vec{p}(t) = \sum_{k=1}^n \vec{u}_k \vec{v}_k^\dagger e^{\nu_k t} \vec{p}(0).$$

If Eq. (22) has a stationary distribution in the $t \rightarrow \infty$ limit, then this must correspond to the eigenvalue $\nu_1 = 0$ (let us assume they are ordered). So

$\frac{d}{dt}\vec{p}(\infty) = 0 \implies$ for $\vec{p}(\infty) = \vec{v}_1, \exists \nu_1 = 0 : \underline{Q}\vec{v}_1 = \nu_0\vec{v}_1 = 0$. Therefore, the solution can be written as

$$(24) \quad \vec{p}(t) = \vec{p}(\infty) + \sum_{k=2}^n \vec{u}_k \vec{v}_k^\dagger e^{\nu_k t} \vec{p}(0).$$

The average channel current of Eq. (21) is

$$\langle I(t) \rangle = \gamma(E - v) \sum_{k=1}^n p_k(t) \delta_{1k} = \gamma(E - v) (p_1(\infty) + \sum_{k=2}^n u_{1k} v_{1k} p_1(0) e^{\nu_k t}),$$

which if the chain is stable ($\nu_k < 0 : \forall k > 1$) has steady state

$$\langle I \rangle = \lim_{t \rightarrow \infty} \langle I(t) \rangle = p_1(\infty).$$

The steady state covariance $C(\Delta) = \lim_{t \rightarrow \infty}$ in this case is

$$C_t(\Delta) = \langle I(t)I(t + \Delta) \rangle - \langle I \rangle^2 = \gamma^2(E - v)^2 \sum_{j,k=1}^n \delta_{1j} \delta_{1k} p_j(t) p_k(t + \Delta) - \langle I \rangle^2$$

Hence

$$C_t(\Delta) = p_1(t)p_1(t + \Delta) - \langle I \rangle^2 = \sum_{i,j=2}^n \underbrace{[\vec{u}_i \vec{v}_i^\dagger \vec{v}_j \vec{u}_i^\dagger]_{11}}_{=\delta_{ij}} e^{\nu_i t + \nu_j(t + \Delta)}$$

If $\nu_k < 0 : \forall k > 1$ then

$$(25) \quad C(\Delta) = \sum_{i=2}^n u_{1i} u_{1i} e^{\nu_i \Delta}$$

is a sum of exponentials. The spectral density is a superposition of Lorentzians.

4.1.2 Simulation a Jump process

Do not track individual channels but the channel numbers. Starting in a particular state $\vec{N} = [N_1, \dots, N_n]$ at time t the life time of staying in that state until $t + \tau$ is

$$f(\tau) = \lambda e^{-\lambda \tau}$$

The escape rate of the state is (any reaction occuring)

$$\lambda = \sum_{k=1}^n N_k a_k$$

where a_k are the rate of leaving state k . For example a_3 in the K^+ -channel is

$$a_3 = 2\beta + 2\alpha$$

But which *reaction* did occur? Let j be the reaction (not the state!). For example, there are 8 reactions in the K^+ channel. The probabilities of occurrence associated with any one of them is

$$p(j) = \frac{N_j \zeta_j}{\sum_{k=1}^{N_{\text{reac}}} N_k \zeta_k} = N_j \zeta_j / \lambda$$

In a computer algorithm one can draw the waiting time for the next reaction simply by generating a uniform random number $r_1 \sim U(0, 1)$ and then

$$\tau \leftarrow -\ln(r_1^{-1}) / \lambda$$

The reaction is determined with a second random number $r_2 \sim U(0, 1)$

$$P(j) = \sum_{k=1}^j p(k)$$

$$j \leftarrow \operatorname{argmax}_j P(j) < r_2$$

while

$$\mathbf{r1} = \mathbf{rand}$$

$$\mathbf{r2} = \mathbf{rand}$$

$$\mathbf{tau} = \ln(1/\mathbf{r1}) / \mathbf{a}$$

4.2 Statistically equivalent diffusion process (Orenstein-Uhlenbeck Process)

The jump process discussed in the previous sections is a continuous time Markov-process on a discrete domain, $K(t) \in \mathbb{N}$ with $t \in \mathbb{R}$. A *diffusion process* is a continuous time Markov-process on a continuous domain, $\eta(t) \in \mathbb{R}$ with $t \in \mathbb{R}$.

Can are in search for a a diffusion process such that

$$I = \gamma (p_1(\infty) + \sum_{k=2}^n \eta_k(t)) (E - v)$$

has the same first and second order statistics (in voltage clamp) as Eq. (21)?

Let us try

$$(26) \quad \tau_k(v) \dot{\eta}_k = -\eta_k + \sigma(v) \xi(t) \text{ where } \langle \xi(0) \xi(\Delta) \rangle = \delta(\Delta)$$

To solve it, use the Fourier Transform

$$i\omega\tau\tilde{\eta}(\omega) = -\tilde{\eta}(\omega) + \sigma\chi(\omega)$$

The place holder symbol χ was introduced for the Fourier transform of the stochastic process $\xi(t)$. Rearranging yields

$$\tilde{\eta} = \frac{\sigma\chi(\omega)}{1+i\omega\tau}$$

The spectrum is

$$\tilde{\eta}(\omega)\tilde{\eta}^*(\omega) = \frac{\sigma^2\chi(\omega)\chi^*(\omega)}{1+(\tau\omega)^2}$$

By definition $\chi(\omega)\chi^*(\omega)$ is the Fourier Transform of the covariance function $\delta(\Delta)$ and from Eq. (19) this is one. Hence,

$$\tilde{\eta}(\omega)\tilde{\eta}^*(\omega) = \frac{\sigma^2}{1+(\tau\omega)^2}$$

Applying the inverse Fourier transform results in the correlation function

$$C(t) = \frac{\sigma^2}{\tau} e^{-|t|/\tau}.$$

A super position of independent such OU process

$$\sum_{k=i}^n \eta_i(t)$$

leads to a correlation function with the same structure as in Eq. (25). We identify $\tau_i = 1/\nu_i$ and $\sigma_i = u_{1i}$.

The idea of matching the second order statistics can be formulated in a far more abstract way in terms of the Kramers-Moyal-van-Kampen expansion of the Master equation

$$\dot{p} = \int w(x' \rightarrow x)p(x', t) - w(x \rightarrow x')p(x, t)dx'$$

$$\partial p(x, t) = \sum_{n=1}^{\infty} \frac{\partial^n}{\partial x^n} K_n(x, t)p(x, t)$$

from

5 Information theory for the living

This theory was never ment to be used to describe living systems in which meaning, *i.e.*, the content of a message actually matter. Information theory deals with optimal compression, lossless transmission of signals irrespective of weather it is relevant or a TV program.

Nontheless a look at a quantitative science of communication may be insightfull.

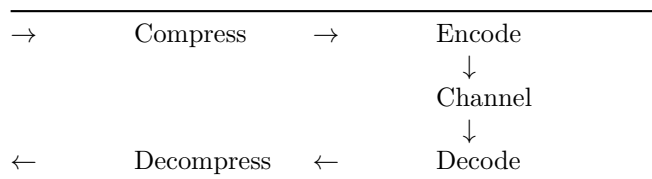
Consult []

Since there is a relation between the PRCs introduced in XXX and the filtering properties of a neuron, one can seek to do the same for phase dynamics and ask what PRC would maximise information transmission. But first, one needs to develop how the PRC predicts a lower bound on the mutual information rate. We begin with a short review of the basic tenets of information theory. Within information theory, a neural pathway is treated as a noisy communication channel in which inputs are transformed to neuronal responses and sent on:

5.1 The communication process

The science of communication is concerned with at least two subtopics: *(i)* the efficient representation of data (compression); and *(ii)* the save transmission of data through unreliable channels.

A source (the message) typically runs through the following processing sequence:



One of the formal assertions of information theory is that these two problems can be addressed separately (without loss of generality or efficiency): Meaning first one compresses by removing redundancies. Then one again adds failsafe redundancy to combat the unreliability.

However convenient for engineering, this does not mean that biological systems have to make use of this fact (source coding and channel coding could be very intertwined).

Also note that many of the mathematical results in information theory are bounds, inequalities not achievable in real systems. To get an idea of the mindset of information theory, check-out *Shannon's source coding theorem*.

5.2 Source coding, data compression and efficient representation

Remember Morse's code. Why does the character *E* only have a single symbol (\cdot), while the character *Q* ($- - \cdot -$) has so many?

The idea could be that: *Highly probable symbols should have the shortest representation, unlikely ones may occupy more space*. What we want to compress is called the *source* and data compression depends on the source distribution. This sounds like a good idea for a biological system: *Do not spend resources on rare events*. Well not quite. Cicadas hear the sounds of their mating partners only once at the very end of a possibly 19 year long life.

International Morse Code

1. The length of a dot is one unit.
2. A dash is three units.
3. The space between parts of the same letter is one unit.
4. The space between letters is three units.
5. The space between words is seven units.

A	• —	U	• • —
B	— • • •	V	• • • —
C	— • — •	W	• — —
D	— • •	X	— • • —
E	•	Y	— • — —
F	• • — •	Z	— — • •
G	— — •		
H	• • • •		
I	• •		
J	• — — —		
K	— • —	1	• — — — —
L	• — • •	2	• • — — —
M	— —	3	• • • — —
N	— •	4	• • • • —
O	— — —	5	• • • • •
P	• — — •	6	— • • • •
Q	— — • —	7	— — • • •
R	• — •	8	— — — • •
S	• • •	9	— — — — •
T	—	0	— — — — —

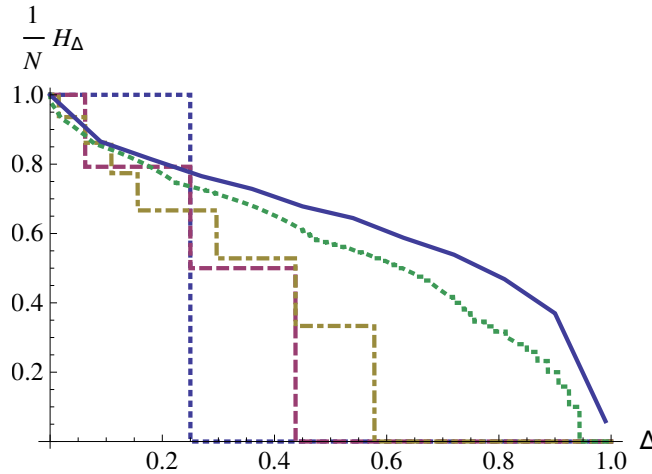
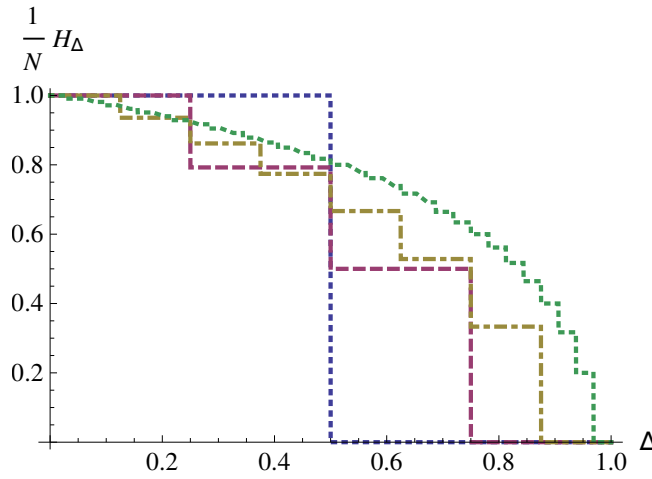
Example (Genetic code): Cells code 20 amino acids with words composed of a four letter⁶ alphabet $A = \{A,G,C,T\}$. Only words of length 3 are sufficient $4^2 = 16 < 20 < 64 = 4^3$. If nature was to require only 16 amino acids, two character words would be sufficient. Only four to drop: Discarding the 4 least probable AS to occur in Proteins: Tryptophan 1.1%, Methionin 1.7%, Histidin 2.1% and Cystein 2.8%. No code words would make for an error of 7.7%.

In general, we can define a risk δ , *i.e.*, the probability of not having a code word for a letter. Then, to to be efficient, one should choose the smallest δ -sufficient subset $S_\delta \subset A$ such that $p(x \in S_\delta) \geq 1 - \delta$.

Def. (Essential bit content): For an alphabet, S_δ , the *essential bit content*, *i.e.*, the number of binary questions asked to identify an element of the set is $H_\delta = \log_2 |S_\delta|$

In the case of the genetic colde the essential bit content for $\delta = 0.077$ is $H_{0.077} = 4$ or if we use the base $b = 4$ for the log it is 2, which is the length of the code words required for 16 AS.

⁶Four nucleobases A: Adenosine, G: Guanine, C: Cytosine, T: Tymine.



Take an even simpler example: We have the alphabet $A = \{0, 1\}$ with probability distributions A: $\{p_0, p_1\} = \{\frac{1}{2}, \frac{1}{2}\}$ and B: $\{p_0, p_1\} = \{\frac{3}{4}, \frac{1}{4}\}$. In figure-?? you can see a plot of $\frac{1}{N}H_\delta$ over the allowed error δ for words of different length N .

For increasing N the curves depend on δ to a lesser degree. Even more it converges to a line around the entropy

$$(27) H = -\sum_k p_k \log_2(p_k)$$

which is $H_A = 1$ and $H_B \approx 0.81$.

This is closely related concept of the *typical set* T^N . Members of the typical set $(x_1, \dots, x_N) = \vec{x} \in T^N$ have

$$-\frac{1}{N} \log p(\vec{x}) \approx H(x).$$

Hence their probability is $p(\vec{x}) \approx 2^{-N H(x)}$, for all of them, which implies the cardinality of the typical set is $2^{N H(x)}$. There holds a kind of law of large

numbers (Asymptotic Equipartition Property, AEP) for the typical set stating that: For N i.i.d. random variables $\vec{x} = (x_1, \dots, x_N)$ with N large \vec{x} is almost certainly a member of T^N ($p(\vec{x} \in T^N) = 1 - \varepsilon$).

Shannon argued that one should therefore base the compression on the typical set and showed that we can achieve a compression down to NH bits.

Asymptotic Equipartition Property (AEP): For i.i.d. random variables

$$-\frac{1}{n} \log p(X_1, \dots, X_n) = -\frac{1}{n} \sum_i \log p(X_i) \rightarrow -\langle \log p(X) \rangle = H(X)$$

Note that the i.i.d. assumption is not too restrictive as we represent correlated processes in terms of the i.i.d. coefficients in their *Karhunen-Loève* transform (or the empirical counter parts).

The typical set $T_n^\varepsilon = \{(x_1, \dots, x_n) : | -1/n \log p(x_1, \dots, x_n) - H(X) | < \varepsilon\}$ allows Shannon's source coding algorithm: The encoder checks if the input sequence lies within the typical set; if yes, it outputs the index of the input sequence within the typical set; if not, the encoder outputs an arbitrary in $n(H + \varepsilon)$ digit number.

By formalising this argument Shannon proved that compression rates up to the source entropy is possible. The converse, that compression below is impossible is a bit more involving.

5.3 Channel coding

Here we consider a „memoryless“ channel:

$$\overset{\text{message}}{W} \xrightarrow{\text{encode}} X \ni x \xrightarrow{\text{noisy channel}} p(y|x) \rightarrow y \in Y \xrightarrow{\text{decode}} \overset{\text{est. Message}}{\hat{W}}$$

The rate R with which information can be transmitted over a channel without loss is measured $\frac{\text{Bits}}{\text{transmission}}$ for a discrete time channel or $\frac{\text{Bits}}{\text{second}}$ for a continuous time channel. Operationally, we wish all bits that are transmitted to be recovered with negligible probability of error.

A measure of information could be:

The average reduction in the number of binary-questions needed to identify $x \in X$ before and after observing $y \in Y$.

This would just be the difference in entropies:

$$M(X; Y) = H(X) - H(X|Y) = - \sum_{x \in X} p(x) \log_2 p(x) + \sum_{\substack{x \in X \\ y \in Y}} p(x, y) \log_2 p(x|y).$$

This goes by the name of *mutual information* or *transinformation*. Remember marginalisation

$$p(x) = \sum_{y \in Y} p(x, y).$$

So the mutual information is

$$M = - \sum_{\substack{x \in X \\ y \in Y}} p(x, y) \log_2 p(x) + \sum_{\substack{x \in X \\ y \in Y}} p(x, y) \log_2 \left(\frac{p(x|y)p(y)}{p(y)} \right)$$

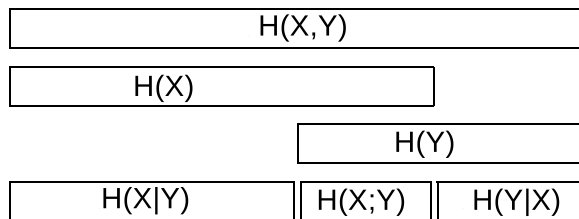
or

$$M(X; Y) = \sum_{\substack{x \in X \\ y \in Y}} p(x, y) \log \frac{p(x, y)}{p(x)p(y)}$$

From the complete symmetry of this quantity we can also write it as

$$M(X; Y) = H(Y) - H(Y|X).$$

The following Figure illustrates how the mutual information is related to respective (conditional) entropies of the input and output ensemble.



We have heard that $I(X; Y)$ quantifies the statistical dependence of X and Y , but how is that related to error-free communication?

$I(X; Y)$ depends on the input ensemble. To focus on the properties of the channel we can simply take an „optimal” input ensemble and define the *channel capacity*

$$C = \max_{p(x)} I(X; Y).$$

It will be left to the sender to actually find the optimal input statistics. Note that $I(X; Y)$ is a concave function (\cap) of $p(x)$ over a convex set of probabilities $\{p(x_i)\}$ (this is relevant for procedures like the Arimoto-Blahut algorithm for estimating C) and hence a local maximum is a global maximum.

```
p = []
while
  p(
```

Shannon established that this capacity indeed measures the maximum amount of error-free information that can be transmitted. A trivial upper bound on the channel capacity is

$$C \leq \min\{\log |X|, \log |Y|\}.$$

This is due to the maximum entropy property of the uniform distribution in the discrete case:

Example (Maximum entropy, discrete case): For the derivate of the entropy from Eq. (27) one gets: $\frac{\partial}{\partial p_i} H = -\log_2 p_i - 1$, which leads to

$p_i \propto e^{-1} \quad \forall i$. After normalisation one has $p_i = 1/N$, so the *uniform distribution* maximises the entropy in the discrete case.

5.4 Information transmission (continuous case)

Information theory has long been in the discussion as an ecological theory upon which to judge the performance of sensory processing (Atick 1992). This led Joseph Atick, Horace Barlow and others to postulate to use this theory to study how nervous systems adapt to the environment. The goal is to make quantitative predictions about what the connectivity in the nervous system and the structure of receptive fields should look like, for instance. In this sense, information theory was hoped to become the basis for an ecological theory of adaptation to environmental statistics (Atick 1992, Barlow (1961)).

Some of the signals in nature (and those applied in the laboratory) are continuous in time and alphabet. Does it make sense to extend the definition of the entropy as

$$(28) \quad H(x) = - \int dx p(x) \ln p(x)?$$

Maybe. Let us see how far one gets with this definition. It is called *differential entropy* by the way. Through quantisation this can be related back to the entropy of discrete alphabets.

If the $p(x)$ is smooth then one associates the probability of being in $i\Delta \leq x \leq (i+1)\Delta$ with

$$p_i = p(x_i)\Delta = \int_{i\Delta}^{(i+1)\Delta} dx p(x)$$

The entropy of the quantised version is

$$\begin{aligned} H_\Delta &= - \sum_{i=-\infty}^{\infty} p_i \ln p_i = - \sum_{i=-\infty}^{\infty} \Delta p(x_i) \ln(p(x_i)\Delta) \\ &= - \sum_{i=-\infty}^{\infty} \Delta p(x_i) \ln p(x_i) - \ln \Delta \end{aligned}$$

This is problematic as the second term goes to infinity for small quantisations. Formally, if $p(x)$ is Riemann integrable, then

$$\lim_{\Delta \rightarrow 0} H_\Delta = H(x) + \ln \Delta = - \int dx p(x) \ln p(x)$$

Since the infinitesimal limit is taken we can also take n to be the number of quantal intervals so that in the limit

$$\lim_{\substack{\Delta \rightarrow 0 \\ n \rightarrow \infty}} \ln \Delta \approx n.$$

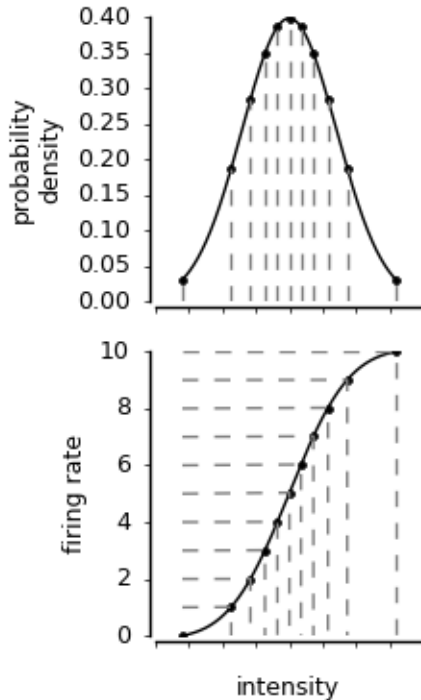
So that an n -bit quantisation of a continuous random variable x has entropy

$$H(x) + n.$$

With the mutual information being the difference of entropies the quantisation term vanishes.

A first example can be taken from rate coding⁷. In terms of our spike trains from Eq. (18), the *instantaneous firing rate* can be defined as

Example (Optimal rate coding in the retina (Laughlin 1981)): In short, Simon measured the distribution of light intensities that would impinge on the fly's compound eye in a natural environment: It was more or less *normally distributed*. He postulated that an efficient allocation of resources (firing rate) would be, to spread more likely inputs over a broader range of firing rates. Why? It makes them easier to differentiate by upstream neurons. Via experiments he found that the fly's compound eye approximates the cumulative probability distribution of contrast levels in natural scenes.



Let us pause and ask: *What is the maximum entropy distribution for a continuous alphabet?*

Example (Maxent, continuous with variance constraint): Take a random variable $x \in \mathcal{R}$. It can not be the uniform distribution as in the discrete case. In fact we need additional constraints. For example one may

⁷Here, the alphabet is a firing rate $f_0 \in \mathcal{R}$. It might be more reasonable to think about spike counts in a given time window, which is still a countable set.

ask for the maximum entropy distribution, $p(x)$, given a fixed mean, μ , and variance σ . Using *Lagrange's multipliers* to formulate the *constraint optimisation* problem

$$C = H + \lambda_0 \int dx p(x) + \lambda_1 (\int dx p(x)x - \mu) + \lambda_2 (\int dx p(x)(x - \mu)^2 - \sigma^2)$$

One searches for $p(x)$ that fulfils $\delta C / \delta p(x) = 0$ and $\partial C / \partial \lambda_i = 0$, where $\delta / \delta p(x)$ is the variational derivative. One finds

$$p(x) = \exp(\lambda_0 - 1 + \lambda_1 x + \lambda_2 (x - \mu)^2).$$

With the normalisation from $\partial C / \partial \lambda_i = 0$ we get the normal distribution ($e^{\lambda_0 - 1} = 1 / \sqrt{2\pi\sigma^2}$, $\lambda_1 = 0$ and $\lambda_2 = 1/\sigma^2$).

What did we learn? Well

$$(29) \quad H(x) \leq H_{\text{Gauss}}(x).$$

If we consider the whole spike train $y(t) = \sum_k \delta(t - t_k^{\text{sp}})$ (see Eq. (18)) as the output, not just its mean input intensity and mean firing rate, we have a continuous time dependent stochastic processes to deal with. Note that the definition of the spike train, if integrated, is related to the *empirical distribution function*.

If we average infinitely many trials we get the instantaneous firing rate $r(t) = \langle y(t) \rangle_{y|x}$. We will give a mathematical definition of $r(t)$ later. Our communication channel looks like that

input signal		neural		response
$x(t) \in \mathbb{R}$	\rightarrow	pathway	\rightarrow	$y(t) \in \mathbb{R}$

The entropy rate of an ergodic process can be defined as the the entropy of the process at a given time conditioned on its past realisations in the limit of large time

$$(30) \quad H[x] = \lim_{t \rightarrow \infty} H[x_t | x_\tau : \tau < t].$$

The mutual information rate measures the amount of information a neural pathway transmits about an input signal $x(t)$ is the mutual information rate,

$$(31) \quad M[x, y] = H[y] - \underbrace{H[y|x]}_{\text{encoding}} = H[x] - \underbrace{H[x|y]}_{\text{decoding}},$$

between the stochastic process, $x(t)$, and the stochastic response process, $y(t)$. The entropy rate H measures the number of discriminable input or output states, either by themselves, or conditioned on other variables.

The mutual information rates, which is the difference between unconditional and conditional entropy rates, characterises the number of input states that can be distinguished upon observing the output. The response entropy rates $H[y]$, for instance, quantifies the number of typical responses per unit time, while $H[x|y]$

is a measure of the decoding noise in the model. If this noise is zero, then the mutual information rate is simply $H[x]$, provided that this is finite.

The conditional entropy rates $H[y|x]$ and $H[x|y]$, characterising the noise in the encoding and decoding model respectively, are each greater than zero. In information theory, these quantities are also called *equivocation*. Hence, both the stimulus and response entropy rates, $H[x]$ and $H[y]$, are upper bounds for the transmitted information.

Example (Info rate continuous discrete-time Gaussian process):

The OU-process from Eq. (26) is an example of a Gaussian process. Take a discretised version, $\vec{x} = [x(t_1), x(t_2), \dots]$, of it such that

$$p(\vec{x}) = |2\pi\underline{K}|^{-1/2} \exp\left(-\frac{1}{2}(\vec{x} - \vec{\mu})^\dagger \underline{K}^{-1}(\vec{x} - \vec{\mu})\right)$$

$$H_n = \frac{1}{2} \ln 2\pi|\underline{K}| + \frac{n}{2}$$

This is similar to the quantisation problem. It might be reasonable to drop the n term (sometimes this is done, sometimes not). For the one dimensional case we have

$$H = \frac{1}{2}(1 + \ln 2\pi\sigma^2)$$

or, if we drop the $n = 1$

$$(32) \quad \sigma^2 = \frac{1}{2\pi} e^{2H_{\text{Gauss}}(x)}$$

Any practical consequences?

Def. (Estimation error): For a random variable x the estimation error of an estimator \hat{x} is

$$\langle (x - \hat{x})^2 \rangle$$

The best estimator is the mean, so the statisticians say. Therefore a lower bound to the estimation error is given by

$$(33) \quad \langle (x - \hat{x})^2 \rangle \geq \langle (x - \langle x \rangle)^2 \rangle = \sigma^2 = \frac{1}{2\pi} e^{2H_{\text{Gauss}}(x)} \geq \frac{1}{2\pi} e^{2H(x)}.$$

The last inequality followed from Eq. (29).

Example (Info rate continuous continuous-time Gaussian process):

Up to an additive constant the entropy of a multivariate Gaussian was^{8,9}

$$H = \frac{1}{2} \ln |\underline{K}| = \frac{1}{2} \text{tr} \ln \underline{K} = \frac{1}{2} \text{tr} \ln \underline{\Lambda} = \frac{1}{2} \sum_k \ln \lambda_k.$$

First let us observe the process for ever $\vec{x} = [x(-\infty), \dots, x(\infty)]$, a bi-infinite series with countable elements. The elements of the covariance matrix $\underline{K}_{ij} = c(i * dt - j * dt)$. The orthogonal eigen-function for the continuous covariance operator on $t \in \mathbb{R}$ are the Fourier bases. It can be shown that in the continuum limit

⁸Remember the determinant is $|\underline{K}| = \prod_{i=1}^n \lambda_i$. So $\ln |\underline{K}| = \sum_{i=1}^n \ln \lambda_i$. In terms of the matrix-logarithm and the trace the determinant can be expressed as $|\underline{K}| = \exp \text{tr} \ln \underline{K}$.

⁹Because the trace is invariant under similarity transforms $\text{tr} \underline{K} = \sum_{i=1}^n \lambda_i$.

$$H = \int df \ln \lambda(f)$$

The result is due to Kolmogorov see also (???,Golshani and Pasha (2010)).

5.5 Linear stimulus reconstruction and a lower bound on the information rate (decoding view)

Without a complete probabilistic description of the model the mutual information can not be calculated. And even with a model the involved integrals may not be tractable. At least two strategies to estimate it exist, though: Either, create a statistical ensemble of inputs and outputs by stimulation, followed by (histogram based) estimation techniques for the mutual information; or, find bounds on the information that can be evaluated more easily. In general, the estimation of mutual information from empirical data is difficult, as the sample size should be much larger than the size of the alphabet. Indeed, each element of the alphabet should be sampled multiple times so that the underlying statistical distribution can, in principle, be accurately estimated. But this prerequisite is often violated, so some techniques of estimating the information from data directly rely on extrapolation (???). The problem becomes particularly hairy when the alphabet is continuous or a temporal processes had to be discretised, resulting in large alphabets.

Another approach, which will allow us to perform a theoretical analysis of phase dynamics, relies on a comparison of the *neuronal “channel”* to the *continuous Gaussian channel* (???,Cpt.~13) is analytically solvable (Cover and Thomas 2012). The approach can be used to estimate the information transmission of neuronal models (???). Also experimental system have ben analysed in this way, *e.g.*:

1. the spider’s stretch receptor (???)
2. the electric sense of weakly electric fish (???) and paddle fish (???)
3. the posterior canal afferents in the turtle (???)

It was proven that in that this method leads to a guaranteed lower bound of the actual information transmitted (???)

If one has experimental control of the stimulus ensemble it can chosen to be a Gaussian process with a flat spectrum up to a cutoff as to not introduce biases for certain frequency bands. The mutual information between stimulus $x(t)$ and response $y(t)$ can be bound from below as

$$(34) \quad M[x, y] = H[x] - H[x|y] \geq H[x] - H_{\text{gauss}}[x|y],$$

Here, $H_{\text{gauss}}[x|y]$ is the equivocation of a process with the same mean and covariance structure as the original decoding noise, but with Gaussian statistics. The conditional entropy of the stimulus given the response is also called recon-

struction noise entropy. It reflects the uncertainty remaining about the stimulus when particular responses have been observed.

It turns out that the inequality in Eq. (33) also holds if the estimator is conditioned. Say from the output of the neuron we estimate its input

$$\hat{x}(t) = \hat{x}_t[y].$$

So if the process has a stationary variance

$$\langle (x(t) - \hat{x}(t))^2 \rangle_{x|y} \geq \inf_{\hat{x}} \langle (x(t) - \hat{x}(t))^2 \rangle_{x|y} = \langle (x(t) - \langle x(t) \rangle_{x|y})^2 \rangle_{x|y} = e^{2H_{\text{gauss}}[x|y]}.$$

The second line uses the fact that in this case the optimal estimator is given by the conditional mean. We have the following bound on the equivocation

$$(35) \quad H[x|y] \leq H_{\text{gauss}}[x|y] \leq \frac{1}{2} \ln \langle (x(t) - \hat{x}(t))^2 \rangle \leq \ln \langle n^2(t) \rangle,$$

The deviation between stimulus and its estimate, $n(t) = x(t) - \hat{x}(t)$, is treated as the noise process.

In order to obtain a tight bound the estimator $\hat{x}(t)$ should be as close to optimal as possible. For the case of additional information given by the response of the neural system $y(t)$ to the process $x(t)$, the estimator should make use of it, $\hat{x}_t[y]$. For simplicity one can assume it is carried out by a filtering operation, $\hat{x}(t) = (f * y)(t)$ specified later (Gabbiani and Koch 1998). Like the whole system the noise process is stationary, and its power spectral density, $P_{nn}(\omega)$, is

$$H_{\text{gauss}}[x|y] \leq \frac{1}{2} \ln \langle n^2(t) \rangle = \frac{1}{2} \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \ln P_{nn}(\omega).$$

Together

$$(36) \quad M[x, y] \geq \frac{1}{2} \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \ln \left(\frac{P_{xx}(\omega)}{P_{nn}(\omega)} \right)$$

So as to render the inequality in Eq. (35) as tight a bound as possible one should use the optimal reconstruction filter from y to \hat{x} . In other words, it is necessary to *extract as much information about x from the spike train y as possible*.

The next step should be to find an expression for the noise spectrum, $P_{nn}(\omega)$, based on the idea of ideal reconstruction of the stimulus. As opposed to the forward filter, the reconstruction filter depends on the stimulus statistics (even without effects such as adaptation). We seek the filter h that minimises the variance of the mean square error

$$(37) \quad \langle n^2(t) \rangle = \langle (x(t) - \hat{x}(t))^2 \rangle, \text{ with } \hat{x}(t) = \int d\tau h(\tau) y(t - \tau).$$

Taking the variational derivative [?] of the error w.r.t.

the filter (coefficients) $h(\tau)$ and equating this to zero one obtains the orthogonality condition for the optimal Wiener filter (???)

$$(38) \quad \langle n(t) y(t - \tau) \rangle = 0, \forall \tau.$$

Inserting the definition of the error, $n(t) = x(t) - \hat{x}(t)$, into Eq. (38) yields

$$\langle x(t) y(t - \tau) \rangle - \int d\tau_1 h(l) \langle r(t - \tau_1) r(t - \tau) \rangle = R_{xy}(\tau) - (h * R_{yy})(\tau) = 0$$

In order to obtain h we need to deconvolve the equation, which amounts to a division in the Fourier domain

$$(39) \quad P_{xy}(\omega) = H(\omega)P_{yy}(\omega) \implies H^{\text{opt}}(\omega) = \frac{P_{xy}(\omega)}{P_{yy}(\omega)}.$$

To compute the mutual information rate, we now calculate the full auto-correlation of the noise when the filter is given by Eq. (39). For an arbitrary filter $h(t)$, we have

$$R_{nn}(\tau) = \langle n(t)n(t+\tau) \rangle = \langle n(t)x(t+\tau) \rangle - \int d\tau_1 h(\tau_1) \langle n(t)y(t+\tau-\tau_1) \rangle.$$

When the orthogonality condition of Eq. (38) holds, the right-most term vanishes. Proceeding by expanding the first term algebraically leads to an expression for the noise correlations

$$R_{nn}(\tau) = \langle n(t)x(t+\tau) \rangle = R_{xx}(\tau) - \int d\tau_1 h(\tau_1)R_{xy}(\tau-\tau_1).$$

This expression can be Fourier transformed in order to obtain the required noise spectrum

$$P_{nn}(\omega) = P_{xx}(\omega) - H(\omega)P_{xy}(\omega) = P_{xx}(\omega) - \frac{|P_{xy}(\omega)|^2}{P_{yy}(\omega)},$$

where the definition of the optimal filter, Eq. (39), was utilised. This result can then be inserted into Eq. (36) to obtain the following well known bound on the information rate (???,lindner2005j:mi,holden1976b,stein1972j:coherenceInfo)

$$(40) \quad \mathcal{M}[x, y] \geq -\frac{1}{2} \int_{-\omega_c}^{\omega_c} \frac{d\omega}{2\pi} \ln \left(1 - \frac{|P_{xy}(\omega)|^2}{P_{xx}(\omega)P_{yy}(\omega)} \right).$$

This information bound involves only spectra and cross-spectra of the communication channel's input and output processes which are experimentally measurable in macroscopic recordings [?, ?, ?, ?]. The channel, in this case the neuron, can remain a black box. But since we can bridge the divide between microscopic, biophysical models and their filtering properties, we will, in the following section, derive the mutual information rates.

Def. (spectral coherence): The expression in Eq. (40) is termed the squared signal response coherence

$$(41) \quad c^2(\omega) = \frac{|P_{xy}(\omega)|^2}{P_{xx}(\omega)P_{yy}(\omega)}.$$

It quantifies the linearity of the relation between x and y in a way that it equals 1 if there is no noise and a linear filter transforms input to output. Both nonlinearities and noise reduce the coherence. The coherence can be estimated from data using the FFT algorithm and spectrum estimation. It is implemented in the free software packages `scipy` and `matplotlib`.

What renders the coherence a useful quantity? While the cross-spectrum informs us when stimulus and output have correlated power in a spectral band, the normalisation with the output auto-spectrum can be crucial. Say we find a

particular high power in $P_{xy}(\omega)$, this may not be related to the stimulus but could just be the intrinsic frequency of the neuron itself.

The coherence is a quantity without mention of the explicit decoding filter, in fact it is symmetric in input and output just as the mutual information. This is beneficial because one can now take the encoding view in the next chapter.

6 Linear response filter

The stimulus spectral density is given by the environment or controlled by the experimental setup, while cross- and output spectra need to be measured or calculated from the model in question. In this lecture cross-spectral and spike train spectral densities are derived from phase oscillator, see Eq. (17), that are in turn derived from biophysical models. This means we do not treat the channel as a blackbox but assume a particular model.

The first quantity we need to calculate Eq. (41) is the cross-spectrum. On the one hand it is the Fourier of the cross-correlation, on the other it can be written as averages of the Fourier transforms (FT and average are linear operation).

$$(42) P_{yx}(\omega) = \langle \langle \tilde{y}(\omega) \tilde{x}^*(\omega) \rangle_{y|x} \rangle_x = \langle \langle \tilde{y}(\omega) \rangle_{y|x} \tilde{x}^*(\omega) \rangle_x.$$

What has happened here? The cross-spectrum can be obtained by averaging the Fourier transformed quantities over trials and the stimulus ensemble. The average can be split into the conditional average over trials $\langle \cdot \rangle_{y|x}$, given a fixed, *frozen* stimulus and the average over the stimulus ensemble, $\langle \cdot \rangle_x$. The former is essentially an average over the encoding noise (Chacron, Lindner, and Longtin 2004, Lindner, Chacron, and Longtin (2005)).

Observe that $\langle \tilde{y}(\omega) \rangle_{y|x}$ is Fourier transform of the trial averaged firing rate conditional on a frozen stimulus

$$r(t) = \langle y(t) \rangle_{y|x}.$$

Thus, it is sufficient to derive a filter that maps input $x(t)$ to a firing rate, not an individual spike train.

Def. (forward, encoding filter): Let $g(t)$ be the filter kernel that maps stimulus into instantaneous firing rate

$$(43) r(t) = (g * x)(t) = \int_{-\infty}^t dr g(r) x(t - r)$$

The filter is causal, since it is implemented by a differential equation and the Laplace transform yields

$$(44) R(s) = G(s)X(s),$$

where $G(s)$ denotes the transfer function of the *encoding filter*.

With this definition the cross-spectrum is written as

$$(45) P_{yx}(\omega) = \langle \langle \tilde{y}(\omega) \rangle_y | \tilde{x}^*(\omega) \rangle_x = G(i\omega) P_{xx}(\omega).$$

This shows us that although we are computing the cross-spectrum of stimulus and spike train the response filter $G(i\omega)$ for the instantaneous firing rate suffices. This simple relation reminds us of the fact that the cross-spectrum is not really a second order quantity, but can be exactly determined by linear response theory. The spike train spectrum $P_{yy}(\omega)$, on the other hand, is truly a second order quantity, viz, the Fourier transform of the auto covariance, and can not be related to the linear response filter without further approximations.

6.1 Instantaneous firing rate in the continuum limit

The instantaneous firing rate can be estimated via binning and trial averaging

$4\Delta r(k\Delta):$	0	3	1	0	0	0	0	3	1	0	0	0	1	2	1	0
Trials:	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0
	0	0	1	0	0	0	0	1	0	0	0	0	0	0	1	0
	0	1	0	0	0	0	0	0	1	0	0	0	0	1	0	0
	0	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0

Two equivalent descriptions¹⁰ of Markov dynamics exist

(i) the *path* view,

$$(46) \dot{x} = u(x, t) = f(x) + g(x)\xi(t);$$

(ii) the *ensemble* view

$$\dot{p}(x, t) = ?.$$

For (i) one can simulate many paths of a stochastic differential equation, with different initial conditions and noise realisations. Histograms can provide the ensemble statistics. But it is also possible to find an evolution equation for the whole ensemble.

The relation between the two can be formally established by the

Def. (empirical measure): Given a stochastic path realisation, $x(t)$, from Eq. (46) the empirical measures is

$$(47) \varrho(y, t) = \delta(x(t) - y).$$

With all the lovely properties of a δ -function.

The probability density

$$(48) p(y, t) = \langle \varrho(y, t) \rangle,$$

¹⁰Similar to the particle and wave duality in quantum mechanics.

where the average is over all paths, $x(t)$, and therefore over all realisations of the noise process $\xi(t)$.

The chain rule yields

$$\frac{\partial}{\partial t} \varrho(y, t) = \dot{x}(t) \frac{\partial}{\partial x} \varrho(y, t) = -\frac{\partial}{\partial y} \varrho(y, t) u(x(t), t)$$

Solving such a PDE involves time integration or other integral transformations (Fourier and Laplace's). Since

$$\int dy \delta(x(t) - y) f(y) = f(x(t)) = \int dy \delta(x(t) - y) f(x(t))$$

Therefore

$$(49) \quad \frac{\partial}{\partial t} \varrho(y, t) = -\frac{\partial}{\partial y} \varrho(y, t) u(y, t) = -\frac{\partial}{\partial y} f(y) \varrho(y, t) - \frac{\partial}{\partial y} g(y) \xi(t) \varrho(y, t)$$

Averaging on both sides results in

$$\frac{\partial}{\partial t} p(y, t) = -\frac{\partial}{\partial y} f(y) p(y, t) - \frac{\partial}{\partial y} g(y) \langle \xi(t) \varrho(y, t) \rangle.$$

The correlation between a stochastic process $\xi(t)$ and a nonlinear functional of it is given by the Novikov-Furutsu-Donsker formula

$$(50) \quad \langle \xi(t) \varrho(y, t) \rangle = -\frac{1}{2} \langle \frac{\delta \varrho}{\delta \xi(t)} \rangle = -\frac{1}{2} \frac{\partial}{\partial y} g(y) p(y, t)$$

All together we have the

Def. (Fokker-Planck equation): The FPE corresponding to Eq. (46) is

$$(51) \quad \frac{\partial}{\partial t} p(y, t) = \frac{1}{2} \frac{\partial}{\partial y} g(y) \frac{\partial}{\partial y} g(y) p(y, t) - \frac{\partial}{\partial y} f(y) p(y, t).$$

This is a diffusion equation. It can be written in the form

$$(52) \quad \frac{\partial}{\partial t} p(y, t) = -\frac{\partial}{\partial y} J(y, t).$$

$$J(y, t) = f(y) p(y, t) - \frac{1}{2} g(y) \frac{\partial}{\partial y} g(y) p(y, t)$$

One needs boundary conditions and initial conditions to solve this PDE.

6.2 Phase flux = firing rate

For convenience rewrite the I/O-equivalent phase oscillator from Eq. (17) as

$$(53) \quad \dot{\phi} = f_0 + z(\phi) x(t) + \sigma \xi(t).$$

Here, as opposed to Eq. (17) $\vec{Z}(\phi) \cdot \vec{\eta}(\phi, t)$ was split into the part that results from the presented stimulus, now denoted $x(t)$, and the part that originated from, *e.g.* intrinsic noise. From Eq. (26) the perturbation vector has

$$\vec{\eta} = \begin{pmatrix} x(t) \\ \sigma_1(v(\phi)) \xi_1(t) \\ \vdots \end{pmatrix}.$$

As long as the intrinsic noise is fast enough compared to the stimulus an averaging can be applied¹¹ to obtain an effective diffusion

$$\sigma^2 = \int_0^1 d\phi \sum_i \sigma_i^2(v_{LC}(\phi)),$$

which enters Eq. (53). The benefit is that the corresponding The Fokker-Planck equation

$$(54) \quad \frac{\partial}{\partial t} p(\phi, t) = \frac{\sigma^2}{2} \frac{\partial^2}{\partial \phi^2} p(\phi, t) - \frac{\partial}{\partial \phi} (f_0 + Z(\phi)x(t))p(\phi, t) = -\frac{\partial}{\partial \phi} J(\phi, t)$$

is tractable in a perturbation expansion. But first, remember what is the goal: Identification of the forward filter $g(t)$ in $r(t) = \int_{-\infty}^t dr g(r)x(t-r)$.

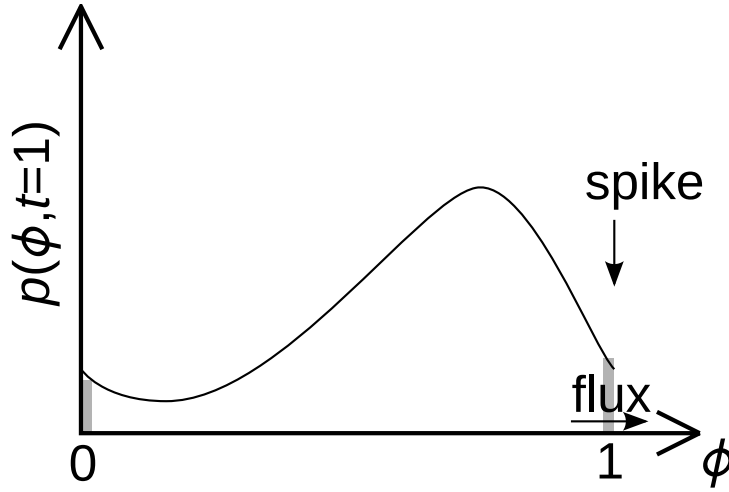


Figure 1: Phase density

$$r(t) = J(1, t) = (f_0 + Z(\phi)x(t))p(\phi, t) - \frac{\sigma^2}{2} \frac{\partial}{\partial \phi} p(\phi, t) \Big|_{\phi=1}$$

The equation is solved with the following perturbation ansatz

$$(55) \quad p(\phi, t) = p_0(\phi) + \sum_i p_i(\phi, t),$$

with the normalisation requirement

$$(56) \quad \int d\phi p_0(\phi) = 1 \text{ and } \forall i \in \mathbb{N}, t \in \mathbb{R} : \int d\phi p_i(\phi, t) = 0.$$

The $p_i(\phi, t)$ are assumed small correction terms, given that the stimulus $x(t)$ is also small.

at the same time one gets

$$J(\phi, t) = J_0(\phi) + \sum_i J_i(\phi, t)$$

¹¹This can be done formally using Chapman-Enskog or adiabatic elimination procedures. The derivation may be included in to future.

In perturbation theory one solves iteratively the equations of the same order

$O(0)$: For the lowest order the stimulus $x(t) = 0$ the FPE is

$$\dot{p}_0 = \frac{\sigma^2}{2} \frac{\partial^2}{\partial \phi^2} p_0 - f_0 \frac{\partial}{\partial \phi} p_0.$$

This equation is homogeneous so we find a time independent steady state solution

$$p_0(\phi) = 1.$$

One may test this by back insertion. Here both the boundary conditions and Eq. (56) are enforced. The solution can be inserted into the definition of the flux to obtain the zeroth order flux

$$J_0 = f_0 p_0 = f_0$$

The next order involves the time-dependent stimulus

$O(x)$: Note that multiplying two terms of order $O(x)$ yields a term of order $O(x^2)$ and is discarded. One is left with

$$\dot{p}_1 = \frac{\sigma^2}{2} \frac{\partial^2}{\partial \phi^2} p_1 - f_0 \frac{\partial}{\partial \phi} p_1 - p_0 x(t) \frac{\partial}{\partial \phi} Z(\phi)$$

To turn the PDE into an algebraic equation one can apply both the Fourier series expansion and the Laplace transform. For this the Laplace transform of the stimulus is denoted as $X(s)$ and the periodic function $Z(\phi) = \sum_{k=-\infty}^{\infty} c_k e^{i2\pi k \phi}$

$$sP_1(k, s) = -\frac{(2\pi k \sigma)^2}{2} P_1(k, s) - f_0 i2\pi k P_1(k, s) - X(s) i2\pi k c_k$$

Solving for $P_1(\phi, s)$

$$P_1(\phi, s) = \sum_{k=-\infty}^{\infty} P_1(k, s) e^{i2\pi k \phi} = \sum_{k=-\infty}^{\infty} \frac{i2\pi k c_k X(s) e^{i2\pi k \phi}}{s + (2\pi k \sigma)^2/2 + i2\pi k f_0}$$

For brevity define the pole $\nu_k = -(2\pi k \sigma)^2/2 - i2\pi k f_0$

The first order flux is

$$\begin{aligned} J_1(k, s) &= f_0 P_1(k, s) + c_k X(s) - \frac{i2\pi k \sigma^2}{2} P_1(k, s) \\ &= f_0 P_1(k, s) + \frac{s - \nu_k}{i2\pi k} P_1(k, s) - \frac{i2\pi k \sigma^2}{2} P_1(k, s) \end{aligned}$$

and

$$i2\pi k J_1(k, s) = i2\pi k f_0 P_1(k, s) + \frac{(2\pi k \sigma)^2}{2} P_1(k, s) + (s - \nu_k) P_1(k, s) = s P_1(k, s)$$

$$J_1(1, s) = \sum_{k=-\infty}^{\infty} \frac{s c_k}{s - \nu_k} X(s)$$

Happily and consistently one finds

$$G(s) = \sum_{k=-\infty}^{\infty} \frac{s c_k}{s - \nu_k}$$

The power spectrum corresponds to the imaginary axis, $G(i\omega)$. The low frequency limit is

$$\lim_{\omega \rightarrow 0} G(i\omega) = c_0 = \langle Z(\phi) \rangle.$$

With $c_k = a_k + ib_k$, the high frequency limit is

$$\lim_{\omega \rightarrow \infty} G(i\omega) = \sum_{k=-\infty}^{\infty} a_k = Z(0).$$

7 Numerical continuation of fixpoints and orbits

In the past lectures functional consequences (*e.g.* filter properties) had been derived from the phase response curve of neurons. The PRCs particular shape (*e.g.* its mean or value at $\phi = 0$) had consequences on what the neuron can do computationally. Next we need to gain some insight into “how a PRC looks like in particular dynamical regimes”. These regimes are reached by changing some system parameter, *e.g.* the membranes leak conductance or capacitance, the composition of ion channels or their kinetic properties.

Often numerical solutions to nonlinear ordinary differential equations are found by (forward) time-integration. An interesting alternative is to track a found solution through parameter space, for which the solution must be *persistent*. If it is *not*, then a *bifurcation* occurs and one observes a *qualitative change* of the solution.

For book on bifurcation theory consult (Izhikevich 2007) and numerical bifurcation analysis (Kielhöfer 2011).

7.1 Continuation of fixed points

Assume for

$$\dot{\vec{x}} = \vec{f}(\vec{x}, p)$$

there is a steady state solution

$$(57) \quad \vec{f}(\vec{x}, p) = 0$$

The fixpoint solution $\vec{x}(p)$ depends on the parameter p . The existence of the solution as a function of the parameter is governed by the *implicit function theorem*.

Implicit function theorem Consider a system of equations

$$\vec{f}(\vec{x}, p) = 0, \text{ with } f \in \mathbb{R}^n, \vec{x} \in \mathbb{R}^n, p \in \mathbb{R} \text{ and } \nabla_{x,p} \vec{f} \in \mathbb{R}^{m \times n+1}.$$

Let f and $\nabla_{x,p} \vec{f}$ be smooth near x . Then if the Jacobian $\nabla_{x,p} \vec{f}$ is nonsingular, \exists a unique, smooth solution family $\vec{x}(p)$ such that

$$\vec{f}(\vec{x}(p), p) = 0.$$

This establishes the existence of lines in a bifurcation diagram.

The method of continuation is a *predictor-corrector* method. In practice, assume the fixpoint is known for one particular parameter value p_0 , then for a small change in parameter the solution is predicted

Predictor step: Taylor's linearisation

$$(58) \quad \vec{x}(p + \delta p) \approx \vec{x}(p) + \delta p \frac{\partial \vec{x}}{\partial p}.$$

To predict the new solution, the old solution is required and the derivative of the solution w.r.t. the parameter that is changed. How to compute the latter? Take the total derivative of Eq. (57) w.r.t. the parameter

$$(\nabla_x f) \frac{\partial \vec{x}}{\partial p} + \frac{\partial \vec{f}}{\partial p} = 0.$$

with formal solution

$$\frac{\partial \vec{x}}{\partial p} = -(\nabla_x \vec{f})^{-1} \frac{\partial \vec{f}}{\partial p}.$$

If $\nabla_x \vec{f}$ is full rank one can use some efficient linear algebra library to find the vector $\frac{\partial \vec{x}}{\partial p}$ and back insert it into Eq. (58). For too large δp the predicted solution will be wrong. Yet, it is a good initial guess form which to find the correct version.

Corrector step: Newton iterations to find the root of $\vec{f}(\vec{x}, p)$

$$(59) \quad \vec{x}_{n+1} = \vec{x}_n - (\nabla_x \vec{f})^{-1} \vec{f}(\vec{x}_n, p)$$

Actually that Newton's iterations are also obtained by linearisation

$$0 = \vec{f}(\vec{x}_{n+1}, p) = \vec{f}(\vec{x}_n) + (\nabla_x \vec{f})(\vec{x}_{n+1} - \vec{x}_n),$$

which if solved for \vec{x}_{n+1} yields Eq. (59).

Convergence analysis of Newton iterations yields that with each newton iterations the number of correct decimal places doubles. Hence often a low number of iterations (3-7) suffice to achieve sufficient numerical accuracy.

Example: Say $f(x, p) = \sqrt{x^2 + p^2} - 1$, then the solution branches are $x(p) = \pm \sqrt{1 - p^2}$. The linear stability analysis $\partial_x f = \frac{x}{\sqrt{x^2 + p^2}}$ that

$$x > 0 \rightarrow \partial_x f > 0 \rightarrow \text{unstable}$$

$$x < 0 \rightarrow \partial_x f < 0 \rightarrow \text{stable}$$

$$\text{Also } \partial_p f = \frac{p}{\sqrt{x^2 + p^2}} \text{ and thus } \partial_p x = \frac{p}{x} = \frac{p}{\pm \sqrt{1 - p^2}}. \text{ What happens at } p = \pm 1 \text{ or } x = 0?$$

Or more general: What happens if we find that the condition number of $\nabla \vec{f}$ explodes?

In the example above the branch switches its stability and it bends back in a *fold* or *turning point*. In general folds can occur with and without stability change.

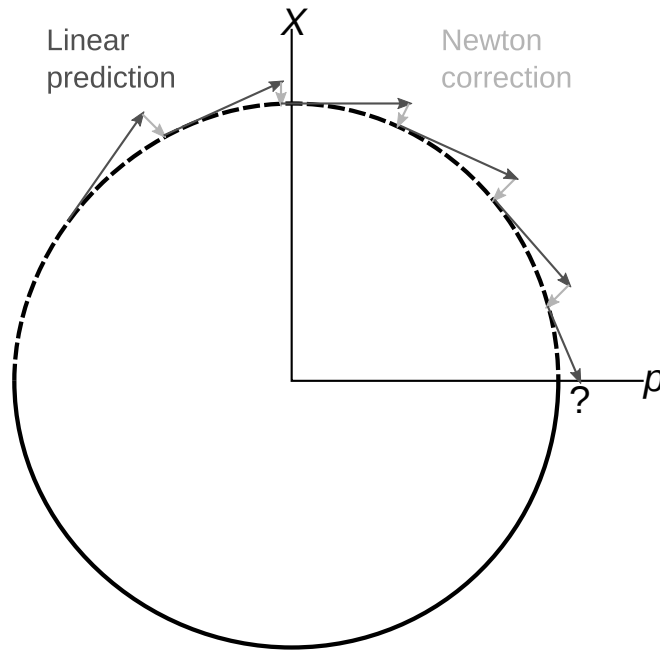


Figure 2: Numerical continuation

7.2 Local bifurcations: What happens if $\nabla \vec{f}$ is singular?

Bifurcating analysis is one big “case discrimination”

7.2.1 Folds and one-dimensional nullspaces

Recall some definitions

Def. (nullspace, kernel): The kernel or nullspace of a matrix \underline{J} is

$$N(\underline{J}) = \{\vec{x} \in \mathbb{R}^n \mid \underline{J}\vec{x} = 0\}$$

Def. (range, image): The range of a matrix \underline{J} is

$$R(\underline{J}) = \{\vec{y} \mid \exists x \in \mathbb{R}^n : \underline{J}\vec{x} = \vec{y}\}$$

Def: Let Q be the projection onto the range

Def: Eigensystem at the fold. Let Jacobian matrix $\underline{J} = \nabla \vec{f}(\vec{x}(0), p(0))$

$$\underline{J}\vec{r}_k = \lambda_k \vec{r}_k \text{ and } \vec{l}_k \underline{J} = \lambda_k \vec{l}_k.$$

So that the nullspace is spanned by \vec{l}_0 .

This section considers $\dim N(\underline{J}) = 1$. Hence, the implicit function theorem is not applicable. From the example above it is apparent that $\vec{x}'(p) = \infty$ at the

bifurcation. The problem can be circumvented by defining a new “arclength parameter”, $p(s)$. The bifurcation branch is then a parametric curve, $(\vec{x}(s), p(s))$. Without loss of generality the bifurcation is to be at $s = 0$.

If the Jacobian matrix $\underline{J} = \nabla \vec{f}$ is rank-deficient the Lyapunov-Schmidt reduction can be applied. Intutiefely the problem is reduced from a high-dimensional, possibly infinite-dimensional, one to one that has as many dimension as the defect of $\nabla \vec{f}$.

The nullspace is spanned by the eigenvector, r_0 of \underline{J} , corresponding to the eigenvalue 0.

Assume that f is twice differentiable w.r.t. \vec{x} , then differentiate $\vec{f}(\vec{x}(s), p(s)) = 0$ w.r.t. s and evaluate at $s = 0$

$$\frac{d}{ds} \vec{f}(\vec{x}(s), p(s)) = \nabla_x \vec{f} \vec{x}'(s) + \partial_p \vec{f} p'(s) = \underline{J} \vec{x}'(s) + \partial_p \vec{f} p'(s)$$

Let $\underline{H} = \nabla \nabla_x \vec{f}$ be the Hessian tensor.

$$\frac{d^2}{ds^2} \vec{f} = \underline{H} \vec{x}''(s) \vec{x}'(s) + \underline{J} \vec{x}''(s) + \partial_p^2 \vec{f} p'(s) + \partial_p \vec{f} p''(s) = 0$$

At $s = 0$ one has $p'(0) = 0$ and $\vec{x}'(0) = \vec{r}_0$. Projecting onto the left-eigenvector \vec{l}_0 to the eigenvalue 0. at $s = 0$ one finds

$$\vec{l}_0 \underline{H} \vec{r}_0 \vec{r}_0 + \vec{l}_0 \partial_p \vec{f} p''(0) = 0$$

or with $\partial_p \vec{f} \notin R(\underline{J})$

$$p''(0) = -\frac{\vec{l}_0 \underline{H} \vec{r}_0 \vec{r}_0}{\vec{l}_0 \partial_p \vec{f}}.$$

This is a test function of wether the bifurcation is

- (i) subcritical ($p''(0) < 0$)
- (ii) supercritical ($p''(0) > 0$)
- (iii) transcritical

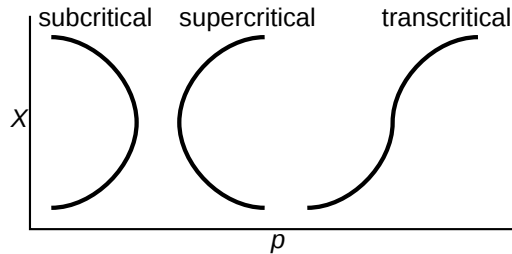


Figure 3: Folds

Def: Let P be the projection onto the null space

There are several cases to be distinguished.

The fold: If the rank deficiency is one-dimensional, $\dim N(\underline{J}) = 1$.

The Andronov-Hopf: If the rank deficiency is two-dimensional.

7.3 Stability exchange

At a simple fold one eigenvalue is zero. Study the eigen system of $\underline{J} = \nabla \vec{f}(\vec{x}(s), p(s))$ near $s = 0$

$$\underline{J}(\vec{r}_0 + \vec{w}(s)) = \lambda(s)(\vec{r}_0 + \vec{w}(s)).$$

With a bit of analysis

$$\lambda'(0) = -\vec{l}_0 \cdot \partial_p \vec{f}(\vec{x}(0), p(0)) p''(0)$$

7.3.1 Extended system

Continue the extended system (linear and nonlinear)

$$\vec{f}(\vec{x}(p), p) = 0$$

$$\nabla \vec{f}(\vec{x}, p) \vec{w} = 0$$

$$(\nabla \vec{f}(\vec{x}, p))^T \vec{z} = 0$$

7.4 Continuation of boundary value problems and periodic orbits

The same procedure than above can be applied to the continuation of periodic orbits and boundary value problems. Define the implicit function to encompass the time derivative

$$\vec{g}(\vec{x}, p) = \frac{d}{dt} \vec{x} - T \vec{f}(\vec{x}, p) = 0, \text{ with } t \in [0, 1].$$

Then proceed as above. Note that the time derivative d/dt is a linear operator which has a matrix representation just like the Jacobian. in that sense

$$\nabla_x \frac{d}{dt} \vec{x} = \frac{d}{dt} \underline{I}$$

8 PRC near the centre manifold

8.1 Dynamics in the centre manifold of a saddle-node

At arbitrary parameters the periodic solution adjoint to the first variational equation on the limit cycle yields the PRC. It can be calculated numerically with the continuation method. Near a bifurcation, however, if major parts of the dynamics happen in the centre manifold the PRC can be calculated analytically.

As an example take the saddle-node on limit cycle bifurcation (SNLC). The spike in this case is a *homoclinic orbit* to a saddle-node, that enters and leaves via the semi-stable (centre) manifold that is associated with the eigenvalue $\lambda_0 = 0$ of the Jacobian at the saddle.

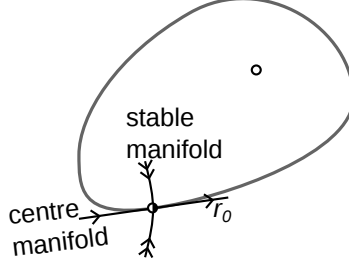


Figure 4: Saddle-node on a limit cycle (SNLC). The dynamics on the stable manifold is fast $\dot{x} = \lambda_1 x$, while the dynamics along the centre subspace is slow $\dot{x} = bx^2$.

$$(60) \quad \dot{\vec{x}} = \vec{f}(\vec{x}) + p\vec{g}(\vec{x})$$

$$\begin{pmatrix} \dot{c}v \\ \tau \dot{n} \end{pmatrix} = \begin{pmatrix} I - I(n, v) \\ n_\infty(v) - n \end{pmatrix}$$

Let there be a saddle-node bifurcation at some value p_0 .

Expanding the right-hand side around *saddle-node fixpoint*, \vec{x}_0 , yields (Ermentrout 1996)

$$\vec{f}(\vec{x}) = \underline{J}(\vec{x} - \vec{x}_0) + \underline{H}(\vec{x} - \vec{x}_0)(\vec{x} - \vec{x}_0) + \dots$$

Establish the eigen system at the saddle-node $\underline{J} = \nabla \vec{f}(\vec{x}_0)$

$$\vec{l}_k \underline{J} = \lambda_k \vec{l}_k \quad \text{and} \quad \underline{J} \vec{r}_k = \lambda_k \vec{r}_k \quad \text{with} \quad \vec{l}_j \cdot \vec{r}_k = \delta_{jk}.$$

By assumption of a saddle-node the Jacobian has a simple zero with an associated eigenvector.

Def (centre subspace): The subspace spanned by \vec{r}_0 is called the *centre subspace* or *slow subspace*.

Write the dynamics around the saddle-node as $\vec{x}(t) = \vec{x}_0 + y\vec{r}_0$. then Eq. (60) is

$$\dot{\vec{x}} + \dot{y}\vec{r}_0 = y\underline{J}\vec{r}_0 + p\vec{g}(\vec{x}_0) + y^2\underline{H}\vec{r}_0\vec{r}_0.$$

Projecting this equation onto the left eigenvector \vec{l}_0 yields the isolated dynamics along the centre manifold:

$$(61) \quad \dot{y} = a + by^2 \quad \text{with} \quad a = p\vec{l}_0 \cdot \vec{g}(\vec{x}_0) \quad \text{and} \quad b = \vec{l}_0 \underline{H} \vec{r}_0 \vec{r}_0.$$

Note: In the literature often y is suggestively written as v assuming the quadratic dynamics is a long the voltage dimension. However, it can be shown that the centre manifold of a conductance-based neuron is never parallel to the voltage dimension.

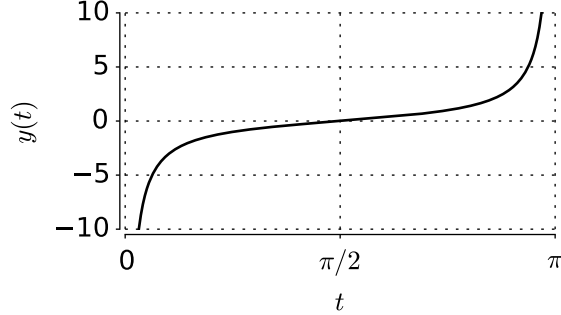


Figure 5: Solution of Eq (61)

The centre manifold is only tangential to the spiking limit cycle dynamics near the saddle. Although the proportion of time spend near the saddle is large at some point the trajectories depart so it is only locally valid.

The formal solution of Eq. (61) with initial condition $y(0) = -\infty$ is

$$(62) \quad y(t) = \sqrt{\frac{a}{b}} \tan(\sqrt{ab}t - \pi/2).$$

Away from the saddle-node the slow manifold accelerates again and the variable y explodes at $\pi/2$. This blowing up is like a spike. So for the SNLC centre manifold one can think of the the spike at $y = \infty$ and reset to $y = -\infty$. The time it takes from $y = -\infty$ to $y = \infty$ is finite and given by

$$T_p = \frac{\pi}{\sqrt{ab}}.$$

Note that for $y(0) = 0$ it is

$$(63) \quad y(t) = \sqrt{\frac{a}{b}} \tan(\sqrt{ab}t)$$

and

$$T_p = \frac{\pi}{2\sqrt{ab}}.$$

The bifurcation parameter enter in a . If $q\vec{g}(\vec{x}) = [c^{-1}I, 0, 0, \dots]^T$, then the firing rate has the typical square-root scaling

$$f_0 = \frac{1}{T_p} = \pi^{-1} \sqrt{bl_{00}c^{-1}(I - I_0)}.$$

Let us check this numerically in the pratical.

8.2 PRC near the centre manifold

PRCs of simple 1D reset-models like quadratic equation of Eq (61) can be calculated. The PRC is defined as the phase gradient (Winfree 2001, Kuramoto

(1984)) with respect to the state variables. In the unperturbed case we have $\dot{\phi} = f_0$, where f_0 is the frequency. The PRC is

$$Z(\phi) = \frac{\partial \phi}{\partial y} = \dot{\phi} \frac{\partial t}{\partial x} = \frac{f_0}{\dot{y}} = \frac{f_0}{a + by^2(f_0 \phi)}.$$

Inserting the solution of Eq. (???) yields

$$\begin{aligned} Z(\phi) &= \frac{f_0}{a + a \tan^2(\pi f_0 \phi - \pi/2)} = a^{-1} f_0 \cos^2(\pi f_0 \phi - \pi/2) \\ &= a^{-1} f_0 \sin^2(\pi f_0 \phi) = \frac{f_0}{2a} (1 - \cos(2\pi f_0 \phi)). \end{aligned}$$

Note that a depends on the bifurcation parameter, yet b does not. Hence it may be preferable to write

$$Z_{\text{SNLC}}(\phi) = \frac{b}{2\pi^2 f_0} (1 - \cos(2\pi f_0 \phi)).$$

A similar calculation for at the saddle-node loop bifurcation yields

$$Z_{\text{SNL}}(\phi) = \frac{b}{2\pi^2 f_0} (1 - \cos(\pi f_0 (1 + \phi))).$$

9 Phase suscetability to channel noise near saddle-node loop bifurcations

The deterministic part of the conductance-based model neuron is

$$\begin{pmatrix} \dot{v} \\ \dot{a}_i \end{pmatrix}$$

The Jacobian is $\underline{J} = \nabla \vec{F}$ has eigen system $\vec{l}_k \underline{J} = \lambda_k \vec{l}_k$ and $\underline{J} \vec{r}_k = \lambda_k \vec{r}_k$, with $\vec{l}_k \cdot \vec{r}_j = \delta_{kj}$.

Proposition 1 Neuron models at a saddle-node on invariant cycle bifurcation, *i.e.*, a simple zero eigenvalue, the centre (slow) manifold is given by

$$\vec{r}_0 = \begin{pmatrix} 1 \\ \frac{d}{dv} a_{\infty, k}(v) \end{pmatrix}.$$

For the usual strictly increasing activation curves, $a'_{\infty, k}(v) > 0$, the centre manifold is not parallel to the voltage or any other state direction.

Proof 1 The special structure of the Jacobian in a conductance-based neuron model, a rank-1 updated diagonal matrix, allows for an explicit calculation of the eigenvector, e_0 , corresponding to the eigenvalue $\lambda_0 = 0$. This gives the direction of the centre manifold. At the stationary point this is $\vec{r}_0 = \begin{pmatrix} 1 \\ a'_{\infty, k}(v) \end{pmatrix}$. As long as the activation functions are strictly monotonous function of the voltage, $a'_{\infty, k}(v) > 0$, the centre manifold has a non-zero component in this gating direction.

The PRC for perturbation along the centre manifold of the SNIC bifurcation is long known. In order to understand perturbations originating from the gating dynamics of different ion channels we need the PRC for arbitrary perturbations.

Proposition 2 The vector of phase-response curves at a SNIC bifurcation is given by

$$\vec{Z}(\phi) = f^{-\frac{1}{2}}(1 - \cos 2\pi\phi)\vec{l}_0.$$

The PRC in all dimensions have the same shape, but their relative scaling differs.

Proof 2 See (Ermentrout and Kopell 1986, Ermentrout (1996), Brown, Moehlis, and Holmes (2004), Ermentrout, Glass, and Oldeman (2012)) for derivation of the phase response curve along the centre manifold of the SNIC. Due to Thm. 1. all dimensions will have the typical quadratic slowing near the saddle-node and can be therefore arbitrarily slower than any other state direction. An perturbation \vec{e}_k along the k^{th} state direction, has according to Thm. 1. always components along the centre manifold $(\vec{l}_0 \cdot \vec{e}_k)\vec{r}_0$. All other directions are exponentially contractive and fast compared to the slow dynamics on the centre manifold. A perturbation \vec{p} can therefore be decomposed into

$$\sum_j (\vec{l}_j \cdot \vec{p}) \vec{r}_j.$$

Along the stable manifolds, $\lambda_j < 0$, all $(\vec{l}_j \cdot \vec{p}) \sim e^{\lambda t}$. Compared to the slow dynamics along the centre manifold the exponential reset back to the limit cycle is instantaneous and, hence, does not affect the phase. It is known that along the centre manifold the quadratic dynamics yields a phase model with a phase-response curve that has the functional form $1 - \cos 2\pi\phi$.

Theorem 3 For both the SNIC and the SNL bifurcations, the peak of the PRC resides at the saddle-node, *i.e.*,

$$\frac{dZ}{d\phi}(\phi)|_{\phi=\phi_{\text{SN}}} = 0.$$

Proof 3 In the limit $f \rightarrow 0$ the isochron density can be made arbitrarily high near the saddle-node. The phase susceptibility in this region will be maximal.

The following holds not only for homoclinic orbits approaching the saddle-node via the semistable centre manifold (the SNIC case), but any homoclinic orbit that approaches via any of the stable manifolds.

The Floquet modes are periodic solution to the first variational system on the limit cycle

$$(64) \quad \vec{W}'_k(\phi) = (\underline{J}(\phi) - \lambda_k \underline{I})\vec{W}_k(\phi) \quad \text{and} \quad \vec{Z}'_k(\phi) = \vec{Z}_k(\phi)(\lambda_k \underline{I} - \underline{J}(\phi))$$

Theorem 4 For homoclinic orbits attached to a saddle-node the tangent plain to the isochrons at the saddle is spanned by the stable manifolds of the saddle, *i.e.*

$$(65) \quad J(\phi_{\text{SN}})W_k(\phi_{\text{SN}}) = \lambda_k W_k(\phi_{\text{SN}}).$$

Hence, $\vec{W}_k(\phi_{\text{SN}})$ is an eigenvector to the Jacobian. The tangent space to the isochrons is thus $T = \{\vec{r}_k : \underline{J}_{\text{SN}}\vec{r}_k = \lambda_k\vec{r}_k \wedge \lambda_k < 0\}$. The PRC is then $\vec{Z}(\phi_{\text{SN}}) \in T^\perp$.

Proof 4 For this one shows that the linearised isochron at $\phi = \phi_{\text{SN}}$ is a solution to the right eigenvalue problem of the Jacobian at the saddle. Since according Thm. 3. the maximum of \vec{Z} resides at $\phi = \phi_{\text{SN}}$ Eq. (65) follows immediately from Eq. (64) with $k = 0$.

10 Event time series and phase descriptions

10.1 Synchronisation and phase locking

If biological oscillators interact their emitted event time series may synchronise. Start with a conductance based models

$$(66) \quad \vec{x}_i = \vec{f}(\vec{x}_i) + \vec{G}(x_i, x_j)$$

and couple them with synapses

$$G(x_i, x_j) = g_{\text{syn}}(v_j)(v_i - E_{\text{syn}}).$$

In Eq (17) the I/O equivalent phase oscillator was derived. Take two phase oscillators i and j that are I/O equivalent to Eq (66)

$$\dot{\phi}_i = f_i + Z(\phi_i)G(\phi_i, \phi_j)$$

$$\dot{\phi}_j = f_j + Z(\phi_j)G(\phi_j, \phi_i)$$

Def (phase locking) Two oscillators i and j are phase locked (a form of *synchronisation*) if their phase difference $\psi_{ij}(t) = \phi_i(t) - \phi_j(t)$ is constant in time. This implies there should be a fixpoint $\dot{\psi} = 0$.

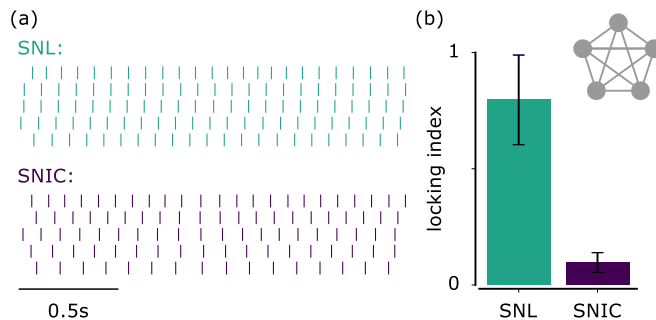


Figure 6: Two small ($n = 5$) all-to-all coupled networks, showing a fixed phase relation (SNL) and no locking (SNIC).

Def (phase locking index) Phase locking can be quantified by evaluating the following temporal average

$$L_{ij} = \langle e^{i(\phi_i(t) - \phi_j(t))} \rangle.$$

The evolution of the phase difference is

$$(67) \quad \dot{\psi} = \Delta f + Z(\phi_j + \psi_{ij})G(\phi_j + \psi_{ij}, \phi_j) - Z(\phi_j)G(\phi_j, \phi_j + \psi_{ij}).$$

Def (frequency detuning) The difference in intrinsic frequencies of the oscillators Δf is called *frequency detuning*.

If the frequency detuning, Δf is small, then ψ is a much slower variable than ϕ_i and ϕ_j . Therefore, The variable ϕ_j in Eq. (67) traverses many periods before ψ changes. In other words: ψ “sees” only the average of ϕ_j . One may apply *averaging theory* to define

$$H(\psi) = \int_0^1 d\phi Z(\phi + \psi)G(\phi + \psi, \phi),$$

then

$$\dot{\psi} = \Delta f + H(\psi) - H(-\psi) = \Delta f + H^{\text{odd}}(\psi).$$

Note that

$$H_{\text{SNLC}}^{\text{odd}}(\psi) = 0$$

and

$$H_{\text{SNL}}^{\text{odd}}(\psi) = (1 - \cos \pi(1 + \psi)) - (1 - \cos(\pi(1 + \psi) + \pi)) = 2 \cos \pi\psi,$$

only the latter has a fixpoint at $\psi = 1/2$

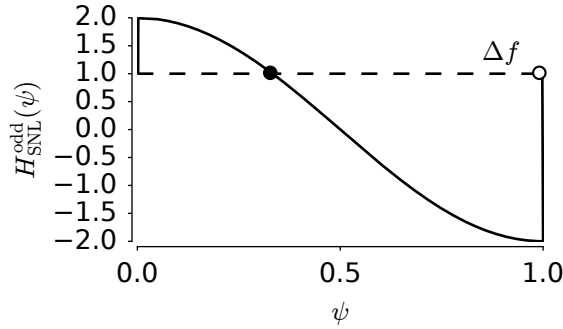


Figure 7: Stable and unstable fixpoints in the SNL coupling function

⇒

- (i) SNLC shows no synchronisation for $\Delta f > 0$
- (ii) SNL shows “antiphase” synchronisation. In a network this may lead to phenomena similar to *frustration*.

10.2 Spike reliability and stimulus entrainment

Not all neuronal populations are coupled. Some sensory neurons like auditory receptors do not have lateral inhibition and just project to the next processing stage. Still many of these neuron receive the same sensory input. Therefore, one can study two neurons that share a common input, or, equivalently one neuron that is stimulated in repetitive trials with the same protocol.

Take two neurons i and j . Each neuron has its own intrinsic noise-level σ_i , but all share and a common stimulus $x(t)$ and mean firing rate f ,

$$(68) \quad \dot{\phi}_i = f + Z(\phi_i)x(t) + \sigma_i \xi_i(t).$$

Remember the neuronal codes. A spike time code was a mapping from a time continuous stimulus $x(t)$ to a spike train $y(t)$, *i.e.*, the output is a high dimensional vector of spike times. In the following the stimulus that is common to all neurons i , is assumed to be a small amplitude zero-mean Gaussian process, $\langle x(t) \rangle = 0$, with wide-sense stationary covariance $C(t-r) = \langle x(t)x(r) \rangle$. The intrinsic noise has different realisation for each neuron.

Q: How *reliable* is this mapping? How *close* in spike-train space are two stimuli? How well is an *idealised* up-stream neuron able to distinguish stimuli, based on the information it receives from the down-stream neurons?

These are decoding questions. They are quantified by looking at *spike-train metrics* (Rossum 2001). In a *non-stationary* environment on other question maybe useful to ask:

Q: Given neurons do lock to a stimulus¹², but that they are in a random state before the stimulus is presented: How long does it take to reliably lock?

This question it important for up-stream neuron, too, since it determines the *minimal window of integration* that is required.

Neurons¹³ are perfectly *inphase*-locking, if their phase difference is zero, $\psi = \phi_i - \phi_j = 0$, and stays that way, $\dot{\psi} = 0$. For simplicity look at the $\sigma = 0$ case. WLOG take ϕ_j as the reference. So, in the present case the phase difference evolves as

$$(69) \quad \dot{\psi} = (Z(\phi_j + \psi) - Z(\phi_j))x(t) = g(\psi, t)$$

In a homogenous (time independent), system information about how fast the locked state is reached and how stable it is is given by the *Lyapunov exponent*, λ of the phase difference.

Def. (Lyapunov exponent): For a deterministic and autonomous¹⁴ system

¹²Not perfectly though, because there is the intrinsic noise, $\xi_i(t)$.

¹³This may now refer to one and the same neuron presented with the same frozen stimulus time and again or a population of non-interacting very similar neurons, which get the same input.

¹⁴independent of t

and a small enough initial phase difference, the Lyapunov exponent is the reciprocal of the exponential attraction or divergence rate $\psi(t) \propto e^{\lambda t}$, which is the solution of the linearised dynamics around the phase fixpoint ψ_0 :

$$\dot{\psi} = \underbrace{g'(\psi_0)}_{\lambda} \psi \text{ or } \lambda = \frac{d}{d\psi} \ln \psi|_{\psi=\psi_0} = g'(\psi_0).$$

Since there a time-continuous stimulus present one can only define an (time) averaged Lyapunov exponent

Def. (average Lyapunov exponent): For a time-dependent system, the averaged Lyapunov exponent is

$$(70) \quad \bar{\lambda} = \left\langle \frac{d}{d\psi} \ln \psi \right\rangle$$

Assume that the neurons are already similar in their phase dynamics, then the right hand side phase difference, Eq. (69) can be expanded around ϕ_j to obtain

$$(71) \quad \dot{\psi} = \psi Z'(\phi_j)x(t).$$

In the absence of intrinsic noise, $\bar{\sigma} = 0$, the averaged Lyapunov exponent from Eq (70) is

$$\bar{\lambda} = \langle Z'(\phi_j)x(t) \rangle.$$

Note that this is a case for the Novikov-Furutsu-Donsker formula, because $Z'(\phi(t))$ is a nonlinear function of a stochastic process, $\phi(t)$ that depends on the stochastic process $x(t)$. Therefore,

$$\bar{\lambda} = \langle Z'(\phi_j)x(t) \rangle = \int_{-\infty}^t dr C(t-r) \left\langle \frac{\delta Z'(\phi_j(t))}{\delta x(r)} \right\rangle.$$

With the chain-rule and the definition in Eq (68) this yields

$$\bar{\lambda} = \int_{-\infty}^t dr C(t-r) \langle Z''(\phi_j(t))Z(\phi_j(r)) \rangle.$$

There are different approaches to calculate the remaining average. In an ergodic system the ensemble average can be replaced by temporal averaging, $\langle \rangle = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T dt$, so one gets

$$\bar{\lambda} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T dt \int_{-\infty}^t dr C(t-r) Z''(\phi_j(t))Z(\phi_j(r)).$$

Further expansion in $x(t)$, *i.e.* to lowest order $\phi_j(t) = f_0 t$, one finds

$$\bar{\lambda} = \lim_{T \rightarrow \infty} \int_0^T dt \int_{-\infty}^t dr C(t-r) Z''(f_0 t)Z(f_0 r).$$

Example (white-noise stimulus) In the special case of a white noise stimulus,

$C(t-r) = \epsilon^2 \delta(t-r)$, one has

$$\bar{\lambda} = \lim_{T \rightarrow \infty} \epsilon^2 \int_0^T dt Z''(f_0 t)Z(f_0 t).$$

Since now we are dealing with periodic functions under the integral this is

$$\bar{\lambda} = f_0 \epsilon^2 \int_0^{f_0^{-1}} dt Z''(f_0 t) Z(f_0 t).$$

In phase variables this is

$$\bar{\lambda} = \epsilon^2 \int_0^1 d\phi Z''(\phi) Z(\phi) = -\epsilon^2 \int_0^1 d\phi (Z'(\phi))^2.$$

In phase locking is guaranteed but the locked state is reached faster if the PRC has large derivatives, *i.e.* higher Fourier components!

Several alternative derivations exist (Teramae and Tanaka 2004, D. S. Goldobin and Pikovsky (2005), Denis S. Goldobin and Pikovsky (2005))

Note (intrinsic noise) In the presence of intrinsic noise ($\sigma > 0$) there is no perfect locking. Nonetheless, the phase difference ϕ may converge to a unimodal distribution, peaked around zero. Given a uniform phase density to start with: *How long does it take to converge to the stationary density of phase differences?*

10.3 Inter-spike statistics

10.3.1 First passage time (no input and constant noise)

In the case $\sigma > 0$ and $x = 0$ we have

$$\dot{\phi} = f_0 + \sigma \xi(t).$$

The adjoint Fokker-Planck equation

$$\dot{p}(\phi, t) = \frac{\sigma^2}{2} \partial_\phi^2 p(\phi, t) + f \partial_\phi p(\phi, t) \text{ s.t. BC } p(1, t) = 1$$

Laplace transform the equation

$$sP(\phi, s) = \frac{\sigma^2}{2} \partial_\phi^2 P(\phi, s) + f \partial_\phi P(\phi, s)$$

A solution is

$$P(\phi, s) = \exp\left\{\frac{f_0}{\sigma^2}(1 - \sqrt{1 + 2s\sigma^2/f_0^2})(1 - \phi)\right\}$$

The inverse Laplace transform of $P(0, s)$ is the inverse Gaussian distribution

$$p(t) = \frac{\exp\left\{-\frac{(tf_0-1)^2}{2\sigma^2 t}\right\}}{\sigma\sqrt{2\pi t}}.$$

Many neuron that do not have slow channels (adaptation, pumps, ...) can be fitted with this ISI distribution.

10.3.2 Phase dependent noise

Ion channel noise is not constant throughout the inter spike interval. In a simple two state channel the voltage dependent noise variance is

$$\sigma^2(v) = \frac{1}{N} n_\infty(v)(1 - n_\infty(v)),$$

where N is the number of ion channels and n_∞ is the steady state activation. Hence one may wish to analyse

$$\dot{\phi} = f_0 + \sigma(\phi)\xi.$$

This can not be solved in general but the moments of the ISI distribution are simpler.

The following concerns equations for the statistics of waiting times as in Ref. (Gardiner 2004). Let $y(t)$ be the process in question, denoting a voltage or a phase variable. The process starts at a particular y_0 , *i.e.* the initial distribution is

$$(72) \quad p(y, t_0) = \delta(y - y_0),$$

and one is interested in the distribution of times for the process to reach $y = y_1$.

If the process is governed by a stochastic differential equation, then it was shown that the associated density $p(y, t)$ is propagated by a specific evolution operator

$$(73) \quad \dot{p}(y, t) = \mathcal{F}(y)p(y, t).$$

This equation was called the Fokker-Planck equation (FPE, see Eq (51)). Denote the solution of a homogeneous FPE with starting distribution concentrated at one value y_0 by $p(y, t; y_0, t_0)$ such that $p(y, t_0; y_0, t_0) = \delta(y - y_0)$ and write its formal solution as

$$(74) \quad p(y, t; y_0, t_0) = e^{(t-t_0)\mathcal{F}(y)}\delta(y - y_0).$$

The goal is to find a relation between the ISI distribution of the neuron model and the FP operator. For that assume the process lives in an interval (y_1, y_2) , where y_2 could denote the spike threshold and y_1 the resting potential to which an IF-like neuron resets, or the two boundaries encapsulating the periodic domain of the phase oscillator interval, *e.g.* $y_1 = 0$ and $y_2 = 1$. At time t_0 , the system is supposed to start inside the interval, $y_1 \leq y_0 \leq y_2$. The probability at time $t > t_0$ of still being inside the interval (y_1, y_2) , and thus no spike occurring, is (Gardiner 2004)

$$G(y_0, t) = \Pr(y_1 \leq y(t) \leq y_2) = \int_{y_1}^{y_2} d\tilde{y} p(\tilde{y}, t; y_0, t_0),$$

with additional condition $G(y_0, t_0) = 1$ because we started with $y_0 \in [y_1, y_2]$. The time derivative of $G(y_0, t)$, *i.e.* the change in the probability of remaining within (y_1, y_2) , at any given t measures the *exit rate* or probability. It is also called

Def (first-passage time density) Thinking of $G(y_0, t)$ as the number density of neurons in $[y_1, y_2]$, the change of this density is the flux out of the interval:

$$(75) \quad q(t, y_0) = \frac{\partial G(y_0, t)}{\partial t}.$$

The goal is to find an evolution equation for $q(t, y_0)$. With the help of the formal solution in Eq. (74), it can be shown that the inner product¹⁵ of $h(y, t) = G(y, -t)$ and $p(y, t; y_0, t_0)$ is constant

$$\begin{aligned} \langle h, p \rangle &= \int dy h(y, t) p(y, t; y_0, t_0) = \iint dy d\tilde{y} p(\tilde{y}, -t; y, t_0) p(y, t; y_0, t_0) \\ &= \iint dy d\tilde{y} e^{-t\mathcal{F}(\tilde{y})} \delta(\tilde{y} - y) e^{t\mathcal{F}(y)} \delta(y - y_0) = \int dy \delta(y - y_0) = 1. \end{aligned}$$

Note that the operator $e^{t\mathcal{F}}$ commutes with the identity operator $\delta(y - \tilde{y})$. Taking the time derivative of this constant and using $\dot{p} = \mathcal{F}p$ one obtains

$$\partial_t \langle h, p \rangle = \langle \dot{h}, p \rangle + \langle h, \dot{p} \rangle = \langle \dot{h}, p \rangle + \langle \mathcal{F}^\dagger h, p \rangle = 0.$$

Because p may change according to its initial conditions, the last expression implies that $\dot{h} = -\mathcal{F}^\dagger h$, or that $G(y, t)$ is a solution to the *adjoint Fokker-Planck* equation (Gardiner 2004)

$$\dot{G}(y, t) = \mathcal{F}^\dagger G(y, t), \text{ s.t. } G(y, T_0) = \mathbb{I}_{[y_1, y_2]}(y).$$

The adjoint operator \mathcal{F}^\dagger is also called the infinitesimal generator of the stochastic process. In addition to the boundary condition above, trivially stating that if we start in the boundary the initial probability of inside is one, one may include reflecting boundary conditions at the lower end $\partial_y G(y, t)|_{y=y_1} = 0$ and absorbing boundary conditions at the upper end $G(y_2, t) = 0$.

Because partial derivatives and integrals are both linear operators, the equation for q directly reads the same

$$(76) \quad \dot{q}(y, t) = \mathcal{F}^\dagger G(y, t),$$

just the boundary condition should read $q(y_2, t) = 1$.

Since one of the main objectives in this document is to establish links between the microscopic noise sources such as channel noise and the macroscopic spike jitter one may immediately pose the question: How much information about the underlying diffusion process can we extract from first passage time densities like the ISI distribution? Might there be a unique diffusion process generating it? A sobering answer to the second question was given in (???): No—the solution is not unique, there are several possible diffusion processes that may lead to one and the same passage time density.

Yet, not all is lost. If one takes into account constraints from membrane biophysics, then the diffusion process derived is not completely arbitrary. In fact, if the model is derived from first principles, then the free parameters in the model can be related to the ISI statistics.

¹⁵The inner product is the inner product on a function space $\langle f, g \rangle = \int dx f(x)g(x)$.

10.3.3 Moments of the ISI distribution

Instead of attempting to obtain the complete ISI distribution by solving the adjoint Fokker-Planck equation, Eq (76) one may content oneself with the first two moments or the coefficient of variation, which one uses to quantify spike jitter in experiments. Let us set $t_0 = 0$ and Denote the n^{th} moment of the ISI distribution as (Gardiner 2004)

$$T_n(y) = \int_0^\infty d\tau \tau^n q(\tau, y) = - \int_0^\infty d\tau \tau^{n-1} G(y, \tau),$$

where the fact was used that for the finite interval $[y_1, y_2]$ exit is guaranteed, *i.e.*, $G(y_0, \infty) = 0$. one may multiply both side of Eq. (76) with t^n and integrate to obtain a recursive ordinary differential equation for the moments

$$(77) \quad nT_{n-1} + \mathcal{F}^\dagger T_n = 0, \text{ s.t. } T'_n(y_1) = T_n(y_2) = 0 \text{ and } T_0 = 1.$$

Here we have imposed reflecting boundary conditions on the lower limit y_1 and absorbing boundary conditions on the upper limit y_2 . These conditions are in agreement with an IF model, which once reaching the spike threshold is reset an unable to move inversely through the spike. As we discussed in the beginning of ?? they can also be applied as an approximation to the phase oscillator if the noise is weak. This equation is also a direct consequence of the Feynman-Kac formula.

In Cpt.~?? the Eq.~(??) will be used to calculate ISI moments of conductance based neurons using a phase reduction. Suppose we have an FP operator $\mathcal{F}(\phi)$ for the equivalent phase variable that is accurate to order ε^k in the noise. Then all moment, T_k , up to order the k^{th} can be obtained accurately. For example if one is interested in ISI variance, the method will require finding a suitable SDEs for the phase variable $\phi(t)$ that gives the FP operator to second order.

10.4 Moments of the inter-spike interval distribution

The n^{th} moment, $T_n(\phi)|_{\phi=1}$, of the first passage time density is the solution to (Gardiner 2004)

$$nT_{n-1} + F(\phi)T_n = 0, \text{ s.t. BC: } T_n(1) = 0 \text{ and } T_0 = 1,$$

where the Fokker-Planck backwards operator for the Stratonovich SDE in Eq ((??)) is

$$F(\phi) = \vec{\sigma}(\phi) \cdot \frac{d}{d\phi} \vec{\sigma}(\phi) \frac{d}{d\phi} + f_0 \frac{d}{d\phi}.$$

Assuming that $\forall \phi : \epsilon(\phi) = f_0^{-1} \vec{\sigma}(\phi) \ll 1$, solutions $T_n(\phi) = T_{n1} + T_{n2} + \dots$ can be sought in a perturbative manner.

10.5 Renewal equation

In a renewal process, all inter-spike intervals are independent, as though each is separately drawn from the ISI distribution. But slow kinetic processes in the neuronal dynamics or long-term correlations in the external stimulus could make the spike train have negative or positive correlations. A point process with such properties would be called a non-renewal.

The ISI distribution alone does not tell us about the correlation between consecutive interspike intervals. Are they independent, negatively or positively correlated? Several types of adaptation currents have time scales spanning orders of magnitude above the spiking period and indeed their contribution to ISI correlations have been analysed [?, ?]. But, for the sake of simplicity, we will ignore the effects on longer time scales and consider a spike train as arising from a renewal process.

In the following we treat the neuron as a threshold device such as an integrate-and-fire neuron or a phase model neuron. We compile here a few known results on renewal processes that we will need in later chapters (*e.g.*, ??).

The transition probability $p(\theta, t; y_0, t_0)$ describes probability a spike occurring at time $t = t_0$, when the neuron was in state y_0 , is followed by a spike at time t , when the neuron crosses threshold θ . For a stationary renewal process, at any given time after a spike the transition probability $p(\theta, t)$ in the renewal case can be decomposed into

$$p(y, t; y_0, 0) = \int_0^\infty d\tau p(y, t; \theta, \tau) q_\theta(\tau) = \int_0^\infty d\tau q_\theta(\tau) p(y, t - \tau; \theta, 0). \quad (1)$$

Here $q_\theta(\tau)$ is shorthand for the interspike interval distribution from y_0 at $t_0 = 0$ to threshold θ , corresponding to the transition probability $p(\theta, \tau; x_0, 0)$. The second equality is due to stationarity, which implies a convolution. The spike autocorrelation $C(\tau)$ is the probability that given a spike at t there is another spike at $t + \tau$. This is equivalent to the transition probability $C(\tau) = p(\theta, \tau; \theta, 0)$ of being back at the spike threshold after τ times has elapsed. By recursively splitting the transition probability in Eq.~(1) into all consecutive possible spiking events one ends with

$$C(\tau) = p(\theta, \tau; \theta, 0) = \sum_{k=0}^{\infty} \underbrace{q_\theta(\tau) * \dots * q_\theta(\tau)}_{k \text{ times}}. \quad (2)$$

The typical approach to isolate the ISI density from Eq.~(1) is by means of Laplace's transform $\tilde{f}(s) = \int_0^\infty dt e^{-st} f(t)$, with $s \in \mathcal{C}$, then

$$\tilde{q}_\theta(s) = \frac{p(x, s; x_0, 0)}{p(x, s; \theta, 0)} \quad (3)$$

In some cases the result may be transformed back into time domain, if the Mellin-Bromwich integral

$$q_\theta(t) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} ds e^{st} \tilde{q}_\theta(s), \quad (4)$$

exists, that is. The constant c is to be chosen to the right of the real parts of all the integrand's singularities. In cases where this integral can not be evaluated explicitly one is stuck with an expression in the Laplace domain, which is not all that bad, as at least individual moments of the time domain distribution as well as the spike power spectrum may be evaluated. Moments are given by

$$\langle \tau^n \rangle = (-1)^n \left. \frac{d^n q_\theta(s)}{ds^n} \right|_{s=0}. \quad (5)$$

The power spectrum of a stationary renewal spike train is the Fourier transform of the spike train autocorrelation Eq.~(2). First, we can identify $C(\tau)$ again in the infinite series and write

$$C(\tau) = q_\theta(\tau) + \int_0^\infty dt q_\theta(t) C(\tau - t). \quad (6)$$

Then, the Lapace transform can be applied to this linear Volterra integral equation to solve for the spiketrain spectrum [?, ?]

$$P(\omega) = r \left\{ 1 + \left. \frac{\tilde{q}_\theta(s)}{1 - \tilde{q}_\theta(s)} \right|_{s \rightarrow i\omega} + \left. \frac{\tilde{q}_\theta(s)}{1 - \tilde{q}_\theta(s)} \right|_{s \rightarrow -i\omega} \right\} \quad (7)$$

10.5.0.1 The first moment ($n = 1$)

The equation for the first moment is

$$(78) \quad 1 + \bar{\sigma}(\phi) \cdot \frac{d}{d\phi} \bar{\sigma}(\phi) \frac{dT_1}{d\phi} + f_0 \frac{dT_1}{d\phi} = 0$$

$\mathcal{O}(\epsilon^0)$ The zeroth order equation reads $1 + f_0 \frac{dT_{10}}{d\phi} = 0$ and the absorbing BC yields

$$T_{10} = f_0^{-1}(1 - \phi).$$

$\mathcal{O}(\epsilon^2)$ Collecting all second order terms from Eq. (78) and inserting T_{10} , results in $\bar{\sigma}(\phi) \cdot \frac{d}{d\phi} \bar{\sigma}(\phi) \frac{dT_{10}}{d\phi} + f_0 \frac{dT_{12}}{d\phi} = f_0 \frac{dT_{12}}{d\phi} - f_0^{-1} \bar{\sigma}(\phi) \cdot \bar{\sigma}'(\phi) = 0$. With the BC this is solved by

$$T_{12}(\phi) = -f_0^{-2} \int_{\psi}^1 d\theta \vec{\sigma}(\theta) \cdot \vec{\sigma}'(\theta) = \frac{1}{2f_0^2} [\sigma^2(\phi) - \sigma^2(1)].$$

10.5.0.2 The second moment ($n = 2$)

The second moment has to solve

$$(79) \quad 2T_1 + \vec{\sigma}(\phi) \frac{d}{d\phi} \vec{\sigma}(\phi) \frac{dT_2}{d\phi} + f_0 \frac{dT_2}{d\phi} = 0$$

$\mathcal{O}(\epsilon^0)$ To zeroth order the equation is $2T_{10} + f_0 \frac{dT_{20}}{d\phi} = 0$, with solution

$$T_{20} = f_0^{-2} (1 - \phi)^2.$$

$\mathcal{O}(\epsilon^2)$ Second order equation is $2T_{12} + \vec{\sigma}(\phi) \cdot \frac{d}{d\phi} \vec{\sigma}(\phi) \frac{dT_{20}}{d\phi} + f_0 \frac{dT_{22}}{d\phi} = 0$, or $3\sigma^2(\phi) - \sigma^2(1) - 2\vec{\sigma}(\phi) \cdot \vec{\sigma}'(\phi)(1 - \phi) + f_0^3 \frac{dT_{22}}{d\phi} = 0$. This is solved by

$$T_{22} = 2f_0^{-3} \int_{\phi}^1 d\theta \sigma^2(\theta) + (\sigma^2(\phi) - \sigma^2(0))(1 - \phi).$$

The ISI variance is given by $T_{22}(0) + T_{20}(0) - (T_{10}(0) + T_{12}(0))^2$, which evaluates to Eq ((???)).

10.6 Spike auto-spectrum

To calculate the spectral coherence, Eq. (41), the spike auto-spectrum $P_{yy}(\omega)$ is still needed for normalisation. In general this is complicated, but if the linearity assumption could be extended to the spike trains itself it helps. Assume

$$y(t) = y_0(t) + \int dr g(r)x(t - r).$$

Then trial averaging, $\langle \cdot \rangle_{y|x}$ yields a result consistent with our previous linear response setting

$$r(t) = r_0 + \int dr g(r)x(t - r).$$

Take $Y_0(\omega) + G(\omega)X(\omega)$ and assume that intrinsic noise and stimulus are uncorrelated

$$\begin{aligned} P_{yy}(\omega) &= \langle (Y_0(\omega) + G(\omega)X(\omega))(Y_0^*(\omega) + G^*(\omega)X^*(\omega)) \rangle \\ &= P_{y_0y_0}(\omega) + |G(\omega)|^2 P_{xx}(\omega) \end{aligned}$$

10.7 Mutual entrainment

Assume two neuron i and j , whose spike-trains are $y(\phi_i(t)) = \sum_k \delta(\phi_i(t) - k)$. The spike dynamics is represented by there I/O-equivalent phase oscillators

$$(80) \quad \phi_i = f_i + Z(\phi_i)y(\phi_j) \text{ and } \phi_j = f_j + Z(\phi_j)y(\phi_i)$$

10.7.1 Spike metric

10.7.2 Time to Spike

11 Appendix

11.1 Novikov-Furutsu-Donsker formula

A relation between Gaussian noise sources and functions of the state variables in stochastic systems is given by the Novikov-Furutsu-Donsker (NFD) formula. It examines the correlation of a stochastic process $\xi(t)$ at a fixed instance in time, t , and a function f of $x(t)$, which is an other stochastic process that depends on $\xi(t)$ (???,(???),(?)). One of the advantages of the NFD formula is that it is applicable to systems with multiplicative noise, as they arise in several applications involving phase response curves. The result is the following

$$(81) \quad \langle f[\xi]\xi(t) \rangle = \frac{1}{2} \int_{-\infty}^{\infty} \langle \xi_t \xi_{t_1} \rangle \left\langle \frac{\delta f[\eta+\xi]}{\delta \eta_{t_1}} \Big|_{\eta=0} \right\rangle dt_1.$$

The script uses the formula several times so a formal and very compact derivation follows (??). In many physical examples only the values in the past $t_1 \leq t$, influence the functional f and the integration range can be adjusted accordingly.

Note that this is related to fluctuation-dissipation relations in statistical physics. The state variable x is a random process that in turn depends on past values of the random process $\xi(t)$. One may, therefore, treat the function f as a functional of the path $\xi_{t_1} : \forall t_1 \leq t$. Assuming that the process $\xi(t)$ has a zero mean function, the first step is to write this functional as a Taylor series¹⁶ around the deterministic function $\eta(t) = 0$ (omitting the integral domain take it to be understood from $-\infty$ to t)

$$f[\eta+\xi] = f[\eta]|_{\eta=0} + \sum_{k=1}^{\infty} \frac{1}{k!} \int \cdots \int dt_1 \cdots dt_k \xi(t_1) \cdots \xi(t_k) \left(\frac{\delta^k f[\eta]}{\delta \eta(t_1) \cdots \delta \eta(t_k)} \right) \Big|_{\eta=0} = \left(\exp \int dt' \xi(t') \frac{\delta}{\delta \eta(t')} \right) f[\eta] \Big|_{\eta=0}.$$

The last expression just a formal, compressed way of writing it using the definition of the exponential displacement operator. As $f[\eta]$ is deterministic it can be yanked from any averaging over the noise ensemble, *e.g.*

$$\langle f[\eta + \xi] \rangle = \left\langle \left(\exp \int dt' \xi(t') \frac{\delta}{\delta \eta(t')} \right) \right\rangle f[\eta] \Big|_{\eta=0}.$$

Hence, we can formally write

$$(82) \quad \langle \xi(t) f[\xi] \rangle = \left\langle \xi(t) \exp \int dt' \xi(t') \frac{\delta}{\delta \eta(t')} \right\rangle f[\eta] \Big|_{\eta=0}$$

¹⁶We are expanding not a function nor a vector-valued function but a functional.

$$= \frac{\langle \xi(t) \exp \int dt' \xi(t') \frac{\delta}{\delta \eta(t')} \rangle}{\langle \exp \int dt' \xi(t') \frac{\delta}{\delta \eta(t')} \rangle} \langle f[\eta + \xi] \rangle_{\eta=0}.$$

Next, the infinite dimensional Fourier transform of a stochastic process called the characteristic functional is introduced

$$\Phi[\lambda] = \langle \exp (i \int dt' \lambda(t') \xi(t')) \rangle.$$

For the, by assumption, Gaussian process ξ_t it is known to be the exponential of a quadratic form

$$\Phi[\lambda] = \exp \left(-\frac{1}{2} \int dt_1 dt_2 \lambda(t_1) C(t_1, t_2) \lambda(t_2) \right),$$

which must be real, $\Phi[\lambda] \in \mathbb{R}$, because the density is symmetric around $\eta(t) = 0$.

With the help of the following identity

$$\frac{\langle \xi(t) \exp i \int dt' \xi(t') \lambda(t') \rangle}{\langle \exp i \int dt' \xi(t') \lambda(t') \rangle} = \frac{\delta}{i \delta \lambda} \ln \langle \exp i \int dt' \xi(t') \lambda(t') \rangle = \frac{\delta}{i \delta \lambda} \ln \Phi[\lambda],$$

and a formal substitution $\delta/\delta \eta(t) \rightarrow i\lambda(t)$ we may simplify Eq.~(??) to

$$\langle \xi_t f[\xi] \rangle = \frac{\delta}{i \delta \lambda} \ln \Phi[\lambda] \langle f[\eta + \xi] \rangle_{\eta=0} = \frac{i}{2} \int dt_1 C(t, t_1) \lambda(t_1) \langle f[\eta + \xi] \rangle_{\eta=0}.$$

Back substituting $i\lambda(t) \rightarrow \delta/\delta \eta_t$ we obtain Eq. (81).

Bibliography

Atick, Joseph J. 1992. “Could Information Theory Provide an Ecological Theory of Sensory Processing?” *Network: Computation in Neural Systems* 3 (2): 213–51. http://www.tandfonline.com/doi/abs/10.1088/0954-898X_3_2_009.

Barlow, HB. 1961. “Possible Principles Underlying the Transformations of Sensory Messages.” In *Sensory Communication*, edited by WA Rosenblith, 217–34. MIT Press.

Brown, Eric, Jeff Moehlis, and Philip Holmes. 2004. “On the Phase Reduction and Response Dynamics of Neural Oscillator Populations.” *Neural Computation* 16 (4): 673–715. doi:10.1162/089976604322860668.

Chacron, Maurice J., Benjamin Lindner, and André Longtin. 2004. “Noise Shaping by Interval Correlations Increases Information Transfer.” *Physical Review Letters* 92 (8). doi:10.1103/PhysRevLett.92.080601.

Chicone, Carmen. 2006. *Ordinary Differential Equations with Applications*. Springer Science & Business Media.

Cover, Thomas M., and Joy A. Thomas. 2012. *Elements of Information Theory*.

John Wiley & Sons.

Ermentrout, Bard. 1996. "Type I Membranes, Phase Resetting Curves, and Synchrony." *Neural Computation* 8 (5): 979–1001. <http://www.mitpressjournals.org/doi/abs/10.1162/neco.1996.8.5.979>.

Ermentrout, G. Bard, Leon Glass, and Bart E. Oldeman. 2012. "The Shape of Phase-Resetting Curves in Oscillators with a Saddle Node on an Invariant Circle Bifurcation." *Neural Computation* 24 (12): 3111–25. doi:10.1162/NECO_a_00370.

Ermentrout, G., and N. Kopell. 1986. "Parabolic Bursting in an Excitable System Coupled with a Slow Oscillation." *SIAM Journal on Applied Mathematics* 46 (2): 233–53. doi:10.1137/0146017.

Fox, Ronald F., and Yan-nan Lu. 1994. "Emergent Collective Behavior in Large Numbers of Globally Coupled Independently Stochastic Ion Channels." *Physical Review E* 49 (4): 3421–31. doi:10.1103/PhysRevE.49.3421.

Gabbiani, Fabrizio, and Christof Koch. 1998. "Principles of Spike Train Analysis." *Methods in Neuronal Modeling* 12 (4): 313–60. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.383.7571&rep=rep1&type=pdf>.

Gardiner, C. W. 2004. *Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences*. 3rd ed. Springer Series in Synergetics. Berlin ; New York: Springer-Verlag.

Goldobin, D. S., and A. S. Pikovsky. 2005. "Synchronization of Self-Sustained Oscillators by Common White Noise." *Physica A: Statistical Mechanics and Its Applications*, New Horizons in Stochastic Complexity International Workshop on New Horizons in Stochastic Complexity, 351 (1): 126–32. doi:10.1016/j.physa.2004.12.014.

Goldobin, Denis S., and Arkady Pikovsky. 2005. "Synchronization and Desynchronization of Self-Sustained Oscillators by Common Noise." *Physical Review E* 71 (4). doi:10.1103/PhysRevE.71.045201.

Goldwyn, Joshua H., and Eric Shea-Brown. 2011. "The What and Where of Adding Channel Noise to the Hodgkin-Huxley Equations." *PLoS Comput Biol* 7 (11): e1002247. doi:10.1371/journal.pcbi.1002247.

Goldwyn, Joshua H., Nikita S. Imennov, Michael Famulare, and Eric Shea-Brown. 2011. "Stochastic Differential Equation Models for Ion Channel Noise in Hodgkin-Huxley Neurons." *Physical Review E* 83 (4): 041908. doi:10.1103/PhysRevE.83.041908.

Golshani, Leila, and Einollah Pasha. 2010. "Rényi Entropy Rate for Gaussian Processes." *Information Sciences* 180 (8): 1486–91. doi:10.1016/j.ins.2009.12.012.

Hodgkin, Alan L., and Andrew F. Huxley. 1952. "A Quantitative Description of Membrane Current and Its Application to Conduction and Excitation in Nerve."

- The Journal of Physiology* 117 (4): 500. <http://www.ncbi.nlm.nih.gov/pmc/articles/pmc1392413/>.
- Izhikevich, Eugene M. 2007. *Dynamical Systems in Neuroscience*. MIT Press.
- Kielhöfer, Hansjörg. 2011. *Bifurcation Theory: An Introduction with Applications to Partial Differential Equations*. Springer Science & Business Media.
- Kuramoto, Y. 1984. *Chemical Oscillations, Waves, and Turbulence*. Springer Science & Business Media.
- Laughlin, S. 1981. “A Simple Coding Procedure Enhances a Neuron’s Information Capacity.” *Zeitschrift Fur Naturforschung. Section C, Biosciences* 36 (9-10): 910–12.
- Linaro, Daniele, Marco Storace, and Michele Giugliano. 2011. “Accurate and Fast Simulation of Channel Noise in Conductance-Based Model Neurons by Diffusion Approximation.” *PLoS Comput Biol* 7 (3): e1001102. doi:10.1371/journal.pcbi.1001102.
- Lindner, Benjamin, Maurice J. Chacron, and André Longtin. 2005. “Integrate-and-Fire Neurons with Threshold Noise: A Tractable Model of How Interspike Interval Correlations Affect Neuronal Signal Transmission.” *Physical Review E* 72 (2). doi:10.1103/PhysRevE.72.021911.
- Orio, Patricio, and Daniel Soudry. 2012. “Simple, Fast and Accurate Implementation of the Diffusion Approximation Algorithm for Stochastic Ion Channels with Multiple States.” *PLoS ONE* 7 (5): e36670. doi:10.1371/journal.pone.0036670.
- Pezo, Danilo, Daniel Soudry, and Patricio Orio. 2014. “Diffusion Approximation-Based Simulation of Stochastic Ion Channels: Which Method to Use?” *Frontiers in Computational Neuroscience* 8: 139. doi:10.3389/fncom.2014.00139.
- Rossum, M. C. W. van. 2001. “A Novel Spike Distance.” *Neural Computation* 13 (4): 751–63. doi:10.1162/089976601300014321.
- Teramae, Jun-nosuke, and Dan Tanaka. 2004. “Robustness of the Noise-Induced Phase Synchronization in a General Class of Limit Cycle Oscillators.” *Physical Review Letters* 93 (20): 204103. doi:10.1103/PhysRevLett.93.204103.
- Winfree, Arthur T. 2001. *The Geometry of Biological Time*. Springer Science & Business Media.