Physics 178/278 - David Kleinfeld - Winter 2022

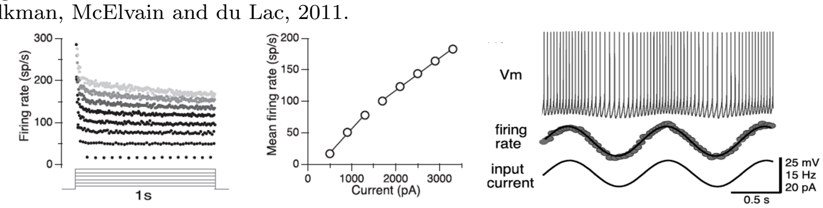
Lecture 8
Revised 26 January 2022 23:05

8 Linear recurrent networks: Integration, line attractors and mono-stability

8.1 Linear recurrent networks and the absence of multistability

We already learned about the role of neuronal thresholds in the ability of networks to store multiple patterns. We now back-up and ask how many stable patterns a linear network can support. This could be the case for networks of cells whose spike rate is uniformly and monotonically modulated up and down (Figure 1). It is also the case for networks of cells without spikes but with solely graded synaptic release. Let's see if we can get a general proof of how many states such networks can support and, if there is something interesting, derive the design rule that relates the desired output to the underlying connectivity.

Figure 1: Precerebellar neurons linearly transform input current into a spiking rate. From Kolkman, McElvain and du Lac, 2011.



To begin, we consider a network with a symmetric weight matrix, $\mathbf{W}$, where as in the past $W_{i,j}$ is the strength of the input to cell i from the output of cell j. The neurons now act as linear devices, *i.e.*, the output of the cell is a linear function of the input and described by a single parameter, the gain, G. Since we are working in the linear regime, we can ignore the difference between cell input, or potential, and firing rate and write the input to the cell as

$$r_i(t) \ = \ \sum_{j=1}^{N} W_{ij} r_j(t) \tag{8.1}$$

where N is the number of neurons and we absorb the gain, G, in the $W_{i,j}$s.

8.1.1 Capacity of a linear network

We start with asking a solely mathematical question. How many states can be stored in a recurrent network with linear interactions? Our work on the "ring" model yielded a linear model that supplied "gain" but led to no storage; that required a nonlinearity. Let's see if there is something more.

We make use of a parallel, clocked updating scheme, in which we explicitly note the time steps, *i.e.*,

$$r_i(t) \;=\; \sum_{j=1}^{N} W_{ij} r_j(t-1). \tag{8.2}$$

In vector notation, this is

$$\vec{\mathbf{r}}(t) = \mathbf{W}\ \vec{\mathbf{r}}(t-1). \tag{8.3}$$

We now iterate, the synchronous equivalent of recurrence, starting from time $t = 0$:

$$\begin{aligned}
\vec{\mathbf{r}}(1) &= \mathbf{W}\ \vec{\mathbf{r}}(0) \qquad (8.4)\\
\vec{\mathbf{r}}(2) &= \mathbf{W}\ \vec{\mathbf{r}}(1)\\
\vec{\mathbf{r}}(3) &= \mathbf{W}\ \vec{\mathbf{r}}(2)\\
&\cdot\\
&\cdot\\
\vec{\mathbf{r}}(n) &= \mathbf{W}\ \vec{\mathbf{r}}(n-1).
\end{aligned}$$

This becomes

$$\vec{\mathbf{r}}(n) = \mathbf{W}^n\ \vec{\mathbf{r}}(0). \tag{8.5}$$

Review of Unitary Transforms

We recall that a matrix $\mathbf{W}$ satisfies an eigenvalue equation

$$\mathbf{W}\ \vec{\mu}_k = \lambda_k \vec{\mu}_k \tag{8.6}$$

where k labels labels the eigenvalue with $k = 1, ..., N$ and includes the case of potential degenerate eigenvectors. The eigenvalues are real numbers when $\mathbf{W}$ is a symmetric matrix whose elements are real. The spectral theorem states that a symmetric matrix whose elements are real can be diagonalized by a matrix transformation by a unity transformation that rotates $\mathbf{W}$ and preserves the eigenvalues, *i.e.*,

$$\mathbf{W} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{\mathbf{T}} \tag{8.7}$$

where $\mathbf{U}$ is a unitary matrix defined through $\mathbf{U}\mathbf{U}^{\mathbf{T}} = \mathbf{I}$ and $\det(\mathbf{U}) = 1$. Each column in $\mathbf{U}$ is one of the eigenvectors $\vec{\mu}_k$, i.e.,

$$\mathbf{U} = \begin{pmatrix} \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \vec{\mu}_1 & \vec{\mu}_2 & \cdots & \vec{\mu}_N \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \end{pmatrix} \quad \text{and} \quad \mathbf{U}^{\mathbf{T}} = \begin{pmatrix} \cdots & \vec{\mu}_1 & \cdots \\ \cdots & \vec{\mu}_2 & \cdots \\ \cdots & \vec{\mu}_3 & \cdots \\ & \cdot & \\ & \cdot & \\ & \cdot & \\ \cdots & \vec{\mu}_N & \cdots \end{pmatrix}$$

and the rotated eigenvectors, $\mathbf{U}^{\mathbf{T}}\vec{\mu}$, are of the form

$$\mathbf{U}^{\mathbf{T}}\vec{\mu}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \end{pmatrix} \qquad \mathbf{U}^{\mathbf{T}}\vec{\mu}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \cdot \\ \cdot \\ \cdot \end{pmatrix} \cdots$$

since $\mathbf{W}\ \vec{\mu}_k = \lambda_k \vec{\mu}_k$ implies $\mathbf{\Lambda}\ \mathbf{U}^{\mathbf{T}}\vec{\mu}_k = \lambda_k \mathbf{U}^{\mathbf{T}}\vec{\mu}_k$, the $\mathbf{U}^{\mathbf{T}}\vec{\mu}_k$ are the eigenvectors of the diagonalized (rotated) system. The diagonal matrix $\mathbf{\Lambda}$ contains the eigenvalues along the diagonal, such that

$$\mathbf{\Lambda} = \begin{pmatrix} \lambda_1 & 0 & 0 & \cdots \\ 0 & \lambda_2 & 0 & \\ 0 & 0 & \lambda_3 & \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \end{pmatrix}$$

We now return to the iterative expression for $\vec{\mathbf{r}}(n)$, i.e.,

$$\begin{aligned} \vec{\mathbf{r}}(n) &= \mathbf{W}^n\ \vec{\mathbf{r}}(0) \\ &= \left(\mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\mathbf{T}}\right)^n\ \vec{\mathbf{r}}(0) \\ &= \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\mathbf{T}}\mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\mathbf{T}} \cdots \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\mathbf{T}}\ \vec{\mathbf{r}}(0) \\ &= \mathbf{U}\mathbf{\Lambda}^{\mathbf{n}}\mathbf{U}^{\mathbf{T}}\ \vec{\mathbf{r}}(0). \end{aligned} \tag{8.8}$$

But the diagonal matrix $\mathbf{\Lambda}^{\mathbf{n}}$, when rank ordered so that λ_1 is the dominant eigenvalue, becomes

$$\mathbf{\Lambda}^n = \begin{pmatrix} \lambda_1^n & 0 & 0 & \cdots \\ 0 & \lambda_2^n & 0 & \\ 0 & 0 & \lambda_3^n & \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \end{pmatrix} = \lambda_1^n \begin{pmatrix} 1 & 0 & 0 & \cdots \\ 0 & \left(\frac{\lambda_2}{\lambda_1}\right)^n & 0 & \\ 0 & 0 & \left(\frac{\lambda_3}{\lambda_1}\right)^n & \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \end{pmatrix} \overrightarrow{n \to \infty}\, \lambda_1^n \begin{pmatrix} 1 & 0 & 0 & \cdots \\ 0 & 0 & 0 & \\ 0 & 0 & 0 & \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \end{pmatrix}.$$

Thus the system converges to a numerical factor times the dominant eigenvector of $\mathbf{W}$, i.e.,

$$\vec{\mathbf{r}}(n)\ \overrightarrow{n \to \infty}\ \lambda_1^n \begin{pmatrix} \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \vec{\mu}_1 & \vec{\mu}_2 & \cdots & \vec{\mu}_N \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & \\ 0 & 0 & 0 & 0 & \\ \cdot & & & & \\ \cdot & & & & \\ \cdot & & & & \\ \cdot & & & & \end{pmatrix} \begin{pmatrix} \cdots & \vec{\mu}_1 & \cdots \\ \cdots & \vec{\mu}_2 & \cdots \\ \cdots & \vec{\mu}_3 & \cdots \\ & \cdot & \\ & \cdot & \\ & \cdot & \\ \cdots & \vec{\mu}_N & \cdots \end{pmatrix} \vec{\mathbf{r}}(0)$$

which becomes

$$\vec{\mathbf{r}}(n)\ \overrightarrow{n \to \infty}\ \lambda_1^n\ [\vec{\mu}_1 \cdot \vec{\mathbf{r}}(0)]\ \vec{\mu}_1 \tag{8.9}$$

and thus only a single state is supported in an iterative network comprised of linear neurons.

The essential issue is neurons that function as linear transducers can support only a single pattern, or stable state. As such, a design rule is to pick a desired stable state, which we will call $\vec{\zeta}$, and compute the weight matrix for our linear network as the outer product

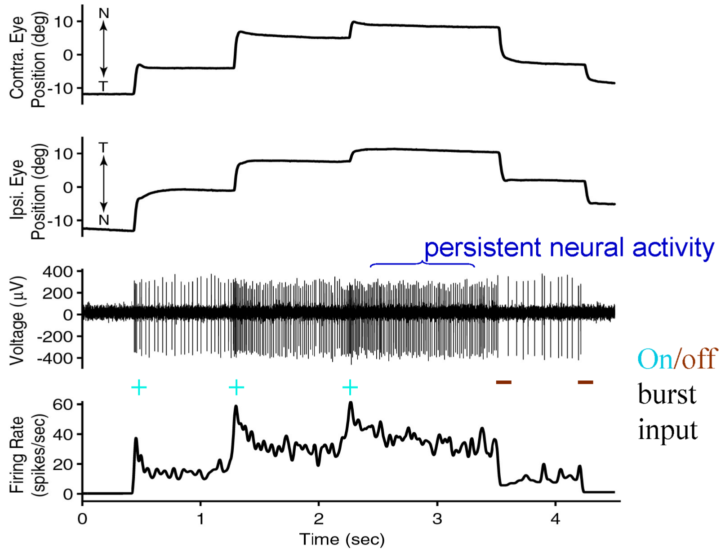
$$\mathbf{W} = \vec{\zeta}\,\vec{\zeta}^T. \tag{8.10}$$

While linear networks will not be useful as associative networks that intrinsically store many patterns, we next show that linear networks can be useful for the particular problem of making a circuit that integrates an input. This comes up for the case of motor systems and was proposed for ocular motor control, first by David Robinson.

8.2 Stability of gaze and the issue of integration

The eyes can remain stable, in the dark and without input, for periods of 20 s in humans . This is nearly two orders of magnitude greater than the time constant of single cells. It is suggestive of a "sample and hold" circuit, or a very low loss integrator. Similar

Figure 2: Multiple levels of stable persistent firing in an oculomotor neural integrator cell in an awake behaving goldfish. Horizontal eye position is measured in the dark. From Major, Baker, Aksay, Mensh, Seung and Tank 2004



behavior is seen in terms of the sustained direction of travel of juvenile zebrafish (Figures 2, 3, and 4).

We will look at a circuit mechanism for such stability, although there are reports of singe cells that function as as integrators of pulses of current (Figure 5).

8.3 Positive feedback and the single neuron

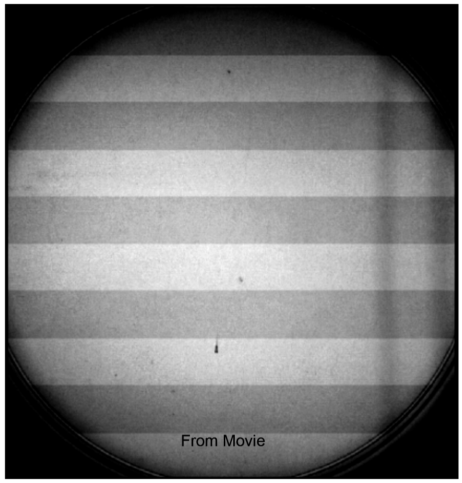
We start with the case of one cell to learn about the importance of integration. Our formalism is in terms of the rate of spiking of the cell. Since we are dealing with linear modeling at this point, we can associate the spike rate with the underlying potential. As such, we write differential equations directly in terms of the rate, which we denote $r(t)$,

$$\tau_0 \frac{dr(t)}{dt} + r(t) = I^{ext}(t) \tag{8.11}$$

where $I^{ext}(t)$ is an external input to the cell normalized in term of rate. This is the same equation for an "RC" circuit in electronics and can be readily solved in terms of a convolution over the input, for which

$$r(t) = r(0)e^{-t/\tau_0} + \int_0^t \frac{dx}{\tau_0}\, e^{-(t-x)/\tau_0} I^{ext}(x). \tag{8.12}$$

Figure 3: Movement in the llarval zebrafish in response to moving stripes. From Misha Ahrens Laboratory



When the input is a constant, *i.e.*, $I^{ext}(t) = I_0^{ext}$, the rate will change toward that constant according to

$$r(t) \;=\; r(0)e^{-t/\tau_0} \;+\; I_0^{ext}(1 - e^{-t/\tau_0}). \tag{8.13}$$

The problem is that this circuit has no memory of the initial rate, $r(0)$ or for that matter the rate at any past time, such as just after a transient input. How can we achieve memory? We consider the addition of positive feedback through an autapse, where the strength of the feedback is set by the scalar constant W. Our rate equation is now

$$\begin{aligned}
\tau_0 \frac{dr(t)}{dt} \;+\; r(t) \;&=\; W r(t) \;+\; I^{ext}(t) \qquad (8.14)\\
\tau_0 \frac{dr(t)}{dt} \;+ (1 \;-\; W)\, r(t) \;&=\; I^{ext}(t)\\
\left(\frac{\tau_0}{1 \;-\; W}\right) \frac{dr(t)}{dt} \;+\; r(t) \;&=\; \frac{I^{ext}(t)}{1 \;-\; W}
\end{aligned}$$

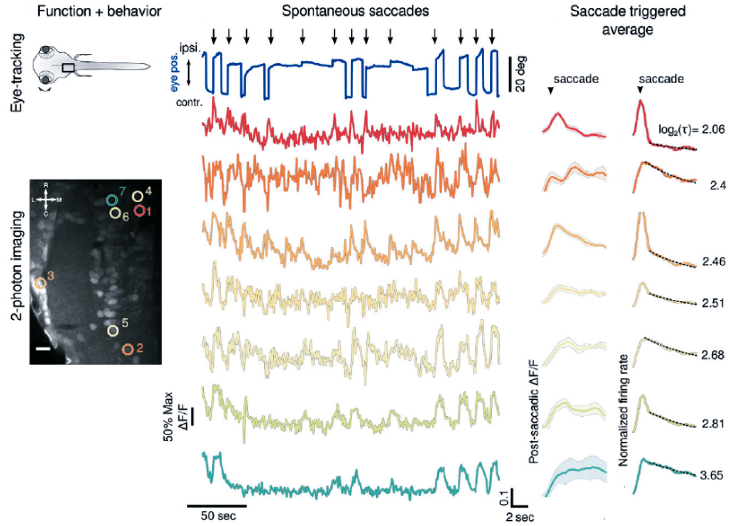
and we see that the time constant is no longer τ_0 but $\frac{\tau_0}{1-W}$. When W approaches a value of $W = 1$ from below, that is, from zero, we see that the effective time constant is very long. In fact, when $W = 1$ it is a perfect integrator with

$$r(t) \;=\; r(0) + \int_0^t \frac{dx}{\tau_0}\; I^{ext}(x). \tag{8.15}$$

When the input is a constant step starting at t=0, the output grows linearly:

$$r(t) \;=\; r(0) + \left(\frac{t}{\tau_0}\right) I_0^{ext}. \tag{8.16}$$

Figure 4: Multiple levels of stable persistent firing in an oculomotor neural integrator cell in a swimming larval zebrafish. Identified integrator neurons are shown in colored circles. From Vishwanathan, Daie, Ramirez, Lichtman, Aksay and Seung, 2002.



Thus if the input is present for only a brief time, say T, the output just shifts from $r(t) = r(0)$ to $r(t) = r(0) + \left(\frac{T}{\tau_0}\right) I_0^{ext}$.

The good news is that we built an integrator - which is a memory circuit - with linear components and positive feedback. The bad news is that W needs to have a value very close to $W = 1$ for the feedback to appreciably extend the time constant. Thus an extension from $\tau_0 = 100$ ms to $\tau = 10$ s, as in the Robinson experiments on the stability of eye position, requires $W = 0.99$. A little variability that causes W to creep up to $W = 1.01$ will lead to an unstable system.

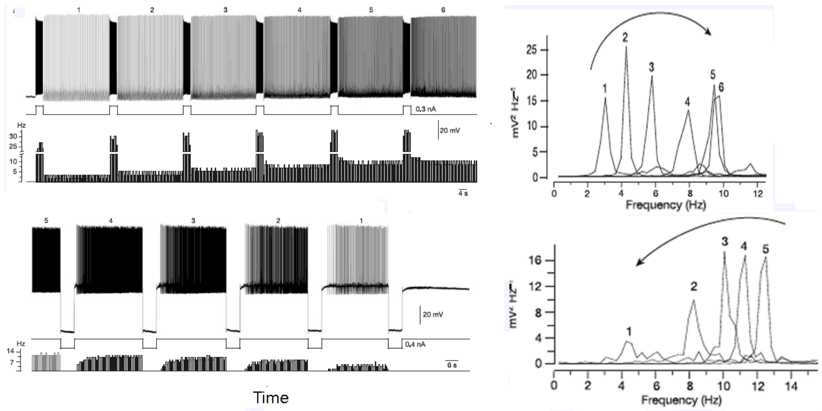
8.4 Stability in a rate based linear network

We learned that a single neuron can function as an integrator. Can we achieve the same behavior in a recurrent linear network, noting that very many neurons are present in biological integrators (Figure 6)? There will be only a single attractor, since linear networks only support one stable state, and we wish to make this the integrator mode. Thus we expect that the path forward will be to transform the set of N coupled linear variables into N uncoupled systems. One of these will be the integrator mode and will must have the largest eigenvalue. Let's see what other constraints arise.

We start with

$$\tau_0 \frac{dr_i(t)}{dt} + r_i(t) = \sum_{j=1}^{N} W_{i,j}\, r_j(t) + I_i^{ext}(t). \tag{8.17}$$

Figure 5: Graded persistent activity in a single cell in entorhinal cortex. Top: Repetitive stimulation with a 4-s depolarizing step gives rise to five distinct increases (traces 1 to 6) of stable discharge. Bottom: Repetitive application of 6-s hyperpolarizing steps gives rise to discrete decreases of stable discharge. From Egorov, Hamam, Fransen, Hasselmo and Alonso, 2002



In vector notion, this becomes

$$\tau_0 \frac{d\vec{\mathbf{r}}(t)}{dt} \ + \ \vec{\mathbf{r}}(t) \ = \mathbf{W}\vec{\mathbf{r}}(t) \ + \ \vec{\mathbf{I}}^{\mathbf{ext}}(t) \tag{8.18}$$

and in equilibrium, for which $\vec{\mathbf{r}}(t) \equiv \vec{\mathbf{r}}_{\mathbf{0}}$,

$$0 = (\mathbf{I} - \mathbf{W}) \ \vec{\mathbf{r}}_{\mathbf{0}} \ - \ \vec{\mathbf{I}}_{\mathbf{0}}^{\mathbf{ext}} \tag{8.19}$$

or

$$\vec{\mathbf{r}}_{\mathbf{0}} \ = \ (\mathbf{I} - \mathbf{W})^{-1} \ \vec{\mathbf{I}}_{\mathbf{0}}^{\mathbf{ext}} \tag{8.20}$$

Is this a stable solution? To address this, we consider a perturbation about $\vec{\mathbf{r}}_{\mathbf{0}}$ and write

$$\vec{\mathbf{r}}(t) \ = \ \vec{\mathbf{r}}_{\mathbf{0}} + \delta\vec{\mathbf{r}}(t) \tag{8.21}$$

Thus

$$0 \ + \ \tau_0 \frac{d\delta\vec{\mathbf{r}}(\mathbf{t})}{dt} \ + \ \vec{\mathbf{r}}_{\mathbf{0}} + \delta\vec{\mathbf{r}}(t) \ = \mathbf{W}\vec{\mathbf{r}}_{\mathbf{0}} \ + \mathbf{W}\delta\vec{\mathbf{r}}(t) \ + \ \vec{\mathbf{I}}_{\mathbf{0}}^{\mathbf{ext}} \tag{8.22}$$

so that the variability about the equilibrium satifies

$$\tau_0 \frac{d\delta\vec{\mathbf{r}}(\mathbf{t})}{dt} \ = - (\mathbf{I} - \mathbf{W}) \ \delta\vec{\mathbf{r}}(t). \tag{8.23}$$

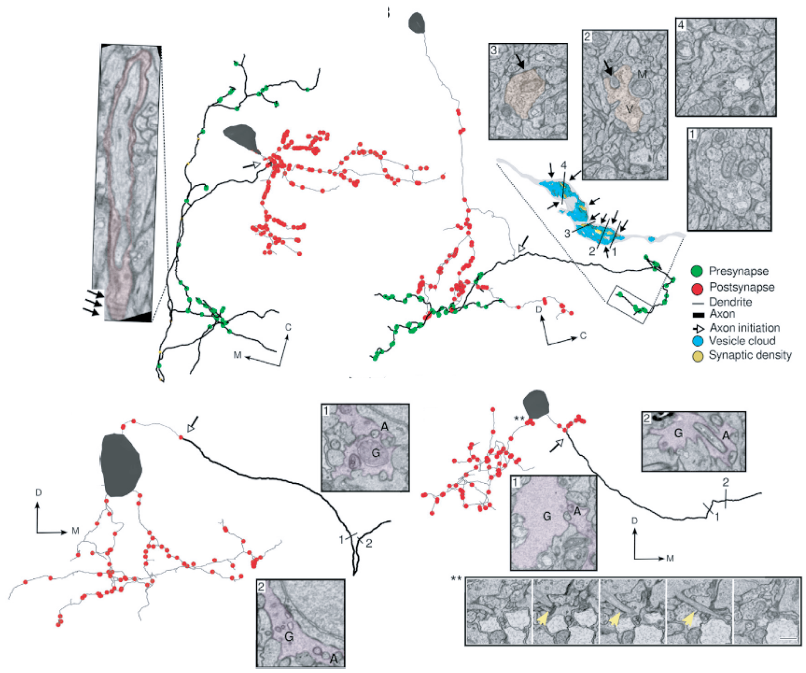
Let us solve this in terms of the eignevectors of $\mathbf{W}$ rather than in terms of the individual $\delta r_i(t)$. In general,

$$\mathbf{W}\vec{\mu}_{\mathbf{i}} = \lambda_i \vec{\mu}_{\mathbf{i}} \tag{8.24}$$

where the $\vec{\mu}_{\mathbf{i}}$ are eigenvectors and the λ_i are the eigenvalues. Then

$$\delta\vec{\mathbf{r}}(t) = \sum_i^N \left[\delta\vec{\mathbf{r}}(\mathbf{t})\right]_i \ \vec{\mu}_{\mathbf{i}} \tag{8.25}$$

Figure 6: Ultrastructural features of integrator neurons with either ipsilaterally (top) or contralateral (bottom) axons. From Vishwanathan, Daie, Ramirez, Lichtman, Aksay and Seung, 2017.



where the expansion coefficients are defined by the overlaps

$$[\delta\vec{\mathbf{r}}(\mathbf{t})]_i \equiv \delta\vec{\mathbf{r}}(t)\cdot\vec{\mu}_i. \tag{8.26}$$

Then

$$\sum_{i=1}^{N}\left(\tau_0\frac{d\,[\delta\vec{\mathbf{r}}(\mathbf{t})]_i}{dt} \;+\; (1-\lambda_i)\,[\delta\vec{\mathbf{r}}(\mathbf{t})]_i\right)\vec{\mu}_{\mathbf{i}} \;=\; 0 \tag{8.27}$$

so that except for the trivial cases $\vec{\mu}_{\mathbf{i}} = 0$ we have

$$\left(\frac{\tau_0}{1-\lambda_i}\right)\frac{d\,[\delta\vec{\mathbf{r}}(\mathbf{t})]_i}{dt} \;+\; [\delta\vec{\mathbf{r}}(\mathbf{t})]_i \;=\; 0 \tag{8.28}$$

for each term. The system is stable if $\lambda_i \leq 1\ \forall i$. The largest eigenvector, taken without loss of generality as λ_1, is the integration mode if it has the largest possible eigenvalue $\lambda_1 = 1$. The other modes will decay away, and suggest the need for $\lambda_i << 1$ for $i \neq 1$.

We now return to the full system and write down a general solution for $\vec{\mathbf{r}}(t)$ in terms of the eigenmodes. Let

$$\vec{\mathbf{r}}(t) = \sum_{i}^{N}[\vec{\mathbf{r}}(\mathbf{t})]_i\ \vec{\mu}_{\mathbf{i}} \tag{8.29}$$

and

$$\vec{\mathbf{I}}^{\mathbf{ext}}(t) = \sum_i^N \left[\vec{\mathbf{I}}^{\mathbf{ext}}(t)\right]_i \vec{\mu}_{\mathbf{i}} \tag{8.30}$$

where $[\vec{\mathbf{r}}(\mathbf{t})]_i \equiv \vec{\mathbf{r}}(t) \cdot \vec{\mu}_i$ and $\left[\vec{\mathbf{I}}^{\mathbf{ext}}(t)\right]_i \equiv \vec{\mathbf{I}}^{\mathbf{ext}}(t) \cdot \vec{\mu}_i$ are time dependent expansion coefficients. Then the original equation of motion

$$\tau_0 \frac{d\vec{\mathbf{r}}(t)}{dt} + \vec{\mathbf{r}}(t) - \mathbf{W}\vec{\mathbf{r}}(t) - \vec{\mathbf{I}}^{\mathbf{ext}}(t) = 0 \tag{8.31}$$

can be written in terms of a differential equation for each eigenmode, *i.e.*,

$$\sum_i^N \left(\tau_0 \frac{d\,[\vec{\mathbf{r}}(\mathbf{t})]_i}{dt} + [\vec{\mathbf{r}}(\mathbf{t})]_i - \lambda_i\,[\vec{\mathbf{r}}(\mathbf{t})]_i - \left[\vec{\mathbf{I}}^{\mathbf{ext}}(t)\right]_i \right) \vec{\mu}_i = 0 \tag{8.32}$$

for which each of the individual terms must go to zero. Thus the effective time constant for the ith mode is

$$\tau_i^{\text{effective}} = \frac{\tau_0}{1-\lambda_i}. \tag{8.33}$$

We can immediately write down the solution for the coefficients for each mode as

$$[\vec{\mathbf{r}}(\mathbf{t})]_i = [\vec{\mathbf{r}}(\mathbf{0})]_i\, e^{-t(1-\lambda_i)/\tau_0} + \int_0^t \frac{dx}{\tau_0}\, e^{-(t-x)(1-\lambda_i)/\tau_0} \left[\vec{\mathbf{I}}^{\mathbf{ext}}(x)\right]_i. \tag{8.34}$$

For the special case of $\lambda_1 = 1$ and $Re\{\lambda_i\} < 1$ for $i > 1$, the dominate mode is also a stable mode, with a firing pattern proportional to $\vec{\mu}_1$, that fulfills our goal and acts as an integrator, *i.e.*,

$$[\vec{\mathbf{r}}(\mathbf{t})]_1 = [\vec{\mathbf{r}}(\mathbf{0})]_1 + \int_0^t \frac{dx}{\tau_0} \left[\vec{\mathbf{I}}^{\mathbf{ext}}(x)\right]_1. \tag{8.35}$$

Thus motion along mode $\vec{\mu}_{\mathbf{1}}$ is driven by the integrated input, with a weight given by the projection. The system is stable when the projected input goes to zero. Motion along other modes will decay away over time as $e^{-(1-\lambda_i)t/\tau_0}$, as the eigenvalues λ_i are less than 1, possibly negative, for $i = 2, 3, ...$ (Figure 7).

8.4.1 Designing the connection matrix

How do we make a connection matrix, $\mathbf{W}$? Our constraint is on the eigenvalues, i.e.,

$$\mathbf{\Lambda} = \begin{pmatrix} 1 & 0 & 0 & \cdots \\ 0 & \lambda_2 & 0 & \\ 0 & 0 & \lambda_3 & \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \end{pmatrix}$$

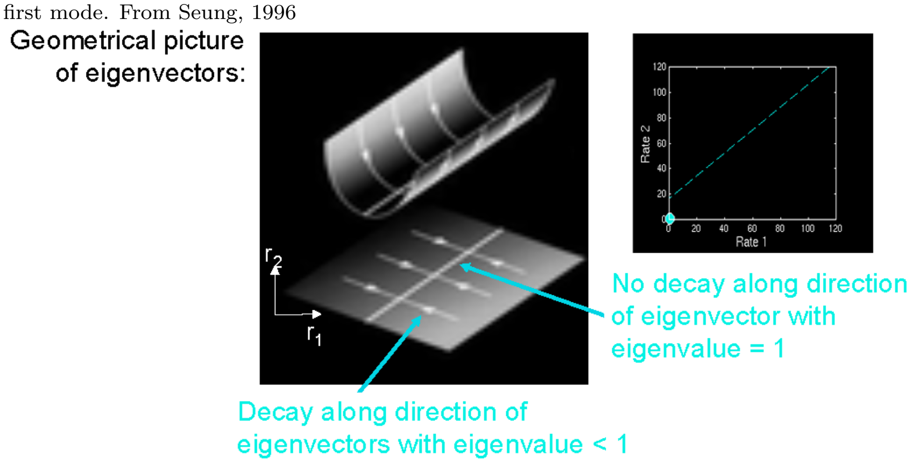


Figure 7: Energy landscape for the line attractor. The first mode in the integration mode, i.e., the line along the valley. All other modes decay in and the system returns to the first mode. From Seung, 1996

where $1 >> \lambda_2 > \lambda_3 > \cdots \lambda_N \geq 0$. As a concrete example, we take a system with two neurons, i.e.,

$$\mathbf{\Lambda} = \begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix}$$

with $1 >> \lambda \geq 0$ and $\mathbf{U}$ as the rotation matrix

$$\mathbf{U} = \begin{pmatrix} \cdot & \cdot \\ \cdot & \cdot \\ \vec{\mu}_1 & \vec{\mu}_2 \\ \cdot & \cdot \\ \cdot & \cdot \end{pmatrix} = \begin{pmatrix} cos\ \eta & sin\ \eta \\ -sin\ \eta & cos\ \eta \end{pmatrix}.$$

with $0 < \eta \ < \pi/2$. Then

$$\mathbf{W} = \mathbf{U\Lambda U^T} = \begin{pmatrix} cos\ \eta & sin\ \eta \\ -sin\ \eta & cos\ \eta \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix} \begin{pmatrix} cos\ \eta & -sin\ \eta \\ sin\ \eta & cos\ \eta \end{pmatrix}$$

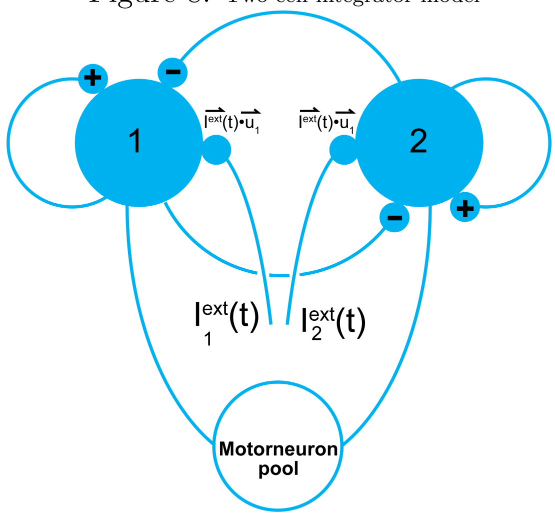
$$= \begin{pmatrix} 1-(1-\lambda)sin^2\ \eta & -\left(\frac{1-\lambda}{2}\right) sin\ 2\eta \\ -\left(\frac{1-\lambda}{2}\right) sin\ 2\eta & 1-(1-\lambda)cos^2\ \eta \end{pmatrix},$$

which is in the form of two neurons that have excitatory self-feedback, as in the single-cell integrator, but mutual inhibition. There are an infinity of such networks since η is a continuous variable, but the signs of the synaptic connections are fixed. As a simple realization, let's take $\eta = \pi/4$, for which

$$\mathbf{W} = \frac{1-\lambda}{2} \begin{pmatrix} +1 & -1 \\ -1 & +1 \end{pmatrix}.$$

In this special case the two neurons are connected by reciprocal inhibition and further enjoy self-excitation (Figure 8). The integrating mode is

Figure 8: Two cell integrator model



$$\vec{\mu}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} +1 \\ -1 \end{pmatrix}$$

and the decaying mode is

$$\vec{\mu}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} +1 \\ +1 \end{pmatrix}.$$

Thus only the differential part of the external input will affect the steady state output of the network.

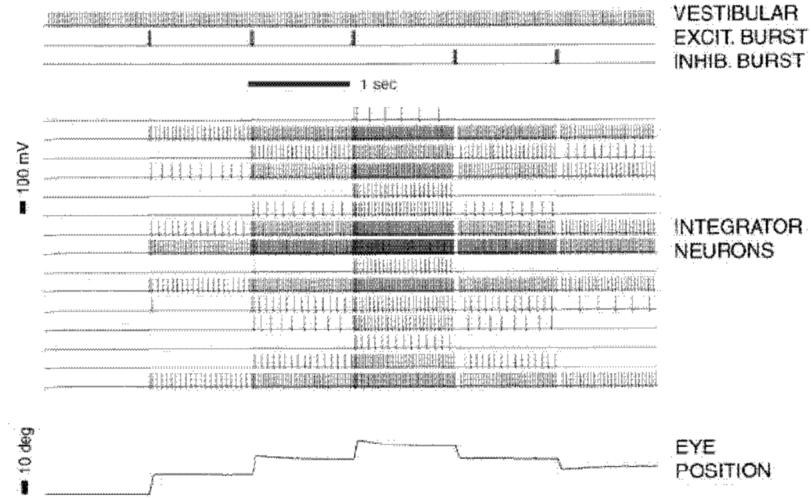
8.4.2 Relation to eye position

As a last step, we relate the neuronal output to eye position. We assume that eye position, denoted $\Theta(t)$, is proportional to a single firing pattern, which makes good sense when that pattern is stable and all others rapidly decay. In fact, this concept makes good sense for any motor act that requires extended stability, such as posture. With reference to angular position, we write

$$\begin{aligned} \Theta(t) &= G \ \vec{\mathbf{r}}(\mathbf{t}) \cdot \vec{\mu}_1 \ + \ \Theta_0 \\ &= G \ [\vec{\mathbf{r}}(\mathbf{t})]_1 + \ \Theta_0 \\ &= G \int_0^t dx \ \left[\vec{\mathbf{I}}^{\mathbf{ext}}(x)\right]_1 \ + \ G \ \left[\vec{\mathbf{I}}_{\mathbf{0}}^{\mathbf{ext}}\right]_1 + \ \Theta_0 \end{aligned} \tag{8.36}$$

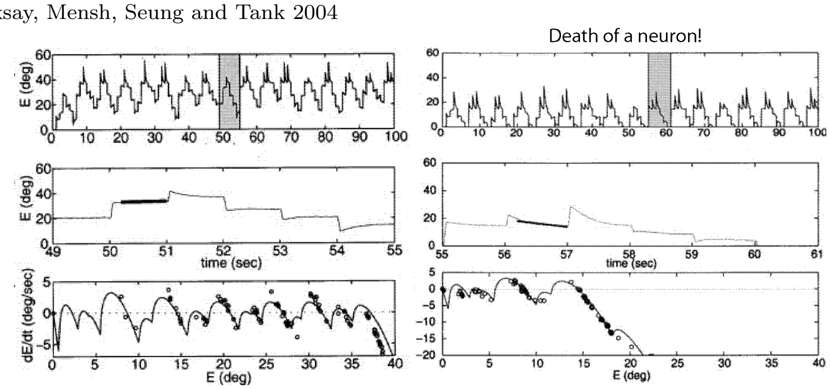
for the amplitude along $\vec{\mu}_1$, where we restore the gain G and denote θ_0 as the baseline position of the eye. The key is that the eye position now follows the integration of the input, or drive, signal (Figure 9).

Figure 9: Simulation of line-attractor model with parameters optimized for the goldfish. From Major, Baker, Aksay, Mensh, Seung and Tank 2004



This model is called a line attractor. The stable output is proportional to a single vector, $\vec{\mu}_1$, but the continuum of points along that vector forms a line in the N-dimensional space of firing rates of the different cells. Changes to $[\vec{\mathbf{r}}(\mathbf{t})]_1$ that result from inputs along the direction of $\vec{\mu}_1$ are along the line. Inputs that are orthogonal to this line have eigenvalues closer to 0 than 1 and rapidly decay so the system returns to the line (Figure 7).

Figure 10: Simulation of line-attractor model with parameters optimized for the goldfish with removal ("death") of one neuron to illustrate sensitivity of stability. From Major, Baker, Aksay, Mensh, Seung and Tank 2004



Experimental tests in which a neuron is removed lead to loss of the integartion mode and are supportive of the model but nonetheless equivocal at this time (Figure 10).