Line attractors

Peter Latham, March 25 2019

Representing continuous variables is easy in a computer – so long as you're willing to live with machine precision. It's harder in networks of neurons, at least on short timescales. That's because activity has to be stable at a continuous (or very finely discretized) set of firing rate. A continuous set of firing rates – the focus here – corresponds to a line attractor: a line in some high dimensional space on which activity is stable. This is the one higher dimensional analog of a point attractor. It doesn't end there, of course; one could imagine n dimensional attractor with n pretty much any integer ($n = 0$ corresponds to point attractors). Our focus here is on constructing a line attractor with networks of neurons, and, more importantly, manipulating the dynamics so that activity can move along the attractor. But much of what I say will be applicable to higher dimensional attractors.

Constructing a line attractor with spiking neurons is possible, but the analysis is pretty hard. So we'll consider instead a network of rate-based neurons. We'll assume they obey the (relatively standard, at least in the rate-based world) equation.

$$\dot{r}_i = \phi\left(\sum_{j=1}^{N} W_{ij} r_j\right) - r_i. \tag{1}$$

Here r_i is the firing rate of neuron i (out of N neurons), a dot denotes a time derivative, and ϕ is the gain function, which is typically approximately sigmoidal. We'll assume that Eq. (1) admits an equilibrium at $r_i = f_i(\theta)$,

$$f_i(\theta) = \phi\left(\sum_j W_{ij} f_j(\theta)\right). \tag{2}$$

What's important is that the equilibrium corresponds to a one dimensional manifold, with position on the manifold determined by θ.

There's an important aside here (it's an aside because we won't use it in our analysis): line attractors are structurally unstable. That means any perturbation to the equations will destroy the line attractor, and turn it into a set of point attractors (and repellers). To see why, write Eq. (2) as

$$f_i(\theta) - \phi\left(\sum_j W_{ij} f_j(\theta)\right) = 0. \tag{3}$$

The left hand side is a function of θ that is exactly zero for all θ. Most functions do not have this property. And any perturbation at all – either to the weights or the function ϕ – will make the left hand side a function of θ, and so destroy the line attractor.

It gets worse: for arbitrary weight matrix W_{ij}, it's extremely hard (and maybe even impossible in general; see the last section) to construct a line attractor. It's a lot easier if W_{ij} is translation invariant: $W_{ij} = W_{i-j}$. In that case, if f_i is an equilibrium of Eq. (1) then so is f_{i+k} for any integer k. This doesn't quite mean you have a line attractor, but it does mean you have a set of point attractors. If N is large they're very closely spaced, which allows you to store continuous variable with precision $1/N$. Moreover, it's not hard to tweak the nonlinearity to give you a true line attractor.

That's the end of the aside; now back to our line attractor. A pure line attractor by itself isn't that interesting: the network just goes to an equilibrium and stays there forever. To do something useful, it must be possible to control the position on the line. For that, though, we need to destroy the line attractor (which is easy to do, as line attractors are structurally unstable), but in controlled ways (which is slightly harder). So we'll add external input. We'll also add recurrent connections that, even without external input, would destroy the line attractor. Such connections will always be there; it's simply not possible for a biological system to admit a perfect line attractor. With these additions, our network equations become

$$\dot{r}_i = \phi\left(\sum_j W_{ij} r_j + \sum_j V_{ij} r_j + h_i\right) - r_i \,. \tag{4}$$

This set of equations is hard to solve in general. To make progress, and to gain insight into its behavior, we'll assume that both V_{ij} and h_i are small. This is reasonable: we want to keep some semblance of our line attractor, and if V_{ij} and h_i are too large, that won't happen.

The smallness assumption suggests a perturbative treatment. We thus let

$$r_i(t) = f_i(\theta_0) + \delta r_i(t) \tag{5}$$

where $f_i(\theta_0)$ is a fixed point of the dynamics given in Eq. (1). We'll eventually choose θ_0 so that $\delta r_i(t)$ is small, and then ask what happens at $t + dt$. Exactly how we do that, which is somewhat important, will be specified below.

Inserting Eq. (5) into (4), Taylor expanding around $r_i = f_i(\theta_0)$, keeping only terms that are linear in δr_i, V_{ij} and h_i (in particular, we throw away the term $\sum_j V_{ij} \delta r_j$), and using Eq. (2), we arrive at

$$\delta\dot{r}_i = \phi_i'(\theta_0)\left[\sum_j W_{ij}\delta r_j + \sum_j V_{ij} f_j(\theta_0) + h_i\right] - \delta r_i \tag{6}$$

where

$$\phi_i'(\theta_0) \equiv \phi'\left(\sum_j W_{ij} f_j(\theta_0)\right) \,. \tag{7}$$

We're using kind of confusing notation: the prime on the right hand side is a derivative; the prime on the left hand side isn't. But it should usually be clear from context what we mean. As usual, I'm using "=" when I should be using "≈". We'll assume that V_{ij} and h_i are small enough that that's OK.

Let us make the definitions

$$J_{ij}(\theta) \equiv \phi_i'(\theta) W_{ij} - \delta_{ij} \tag{8a}$$

$$S_i(\theta) \equiv \phi_i'(\theta) \sum_j V_{ij} f_j(\theta) + \phi_i'(\theta) h_i \tag{8b}$$

where δ_{ij} is the Kronecker delta: it's 1 if $i = j$ and zero otherwise. This allows us to write Eq. (6) in vector notation,

$$\delta\dot{\mathbf{r}} = \mathbf{J}(\theta_0) \cdot \delta\mathbf{r} + \mathbf{S}(\theta_0) \tag{9}$$

where bold denotes vectors and matrices, and "$\cdot$" denotes a dot product: $\mathbf{J} \cdot \delta\mathbf{r} = \sum_j J_{ij}\delta r_j$ (and for two vectors, say $\mathbf{p}$ and $\mathbf{q}$, $\mathbf{p} \cdot \mathbf{q} = \sum_i p_i q_i$).

To solve Eq. (9), we express $\delta\mathbf{r}$ in terms of the eigenvectors of $\mathbf{J}(\theta_0)$. Let the k^{th} eigenvector be $\mathbf{v}_k(\theta_0)$ and its corresponding eigenvalue $\lambda_k(\theta_0)$. We'll also need the adjoint eigenvalues, $\mathbf{v}_k^\dagger(\theta_0)$. These three quantities are defined via the relations (for general θ)

$$\mathbf{J}(\theta) \cdot \mathbf{v}_k(\theta) = \lambda_k(\theta)\mathbf{v}_k(\theta) \tag{10a}$$

$$\mathbf{v}_k^\dagger(\theta) \cdot \mathbf{J}(\theta) = \lambda_k(\theta)\mathbf{v}_k^\dagger(\theta)\,. \tag{10b}$$

The eigenvectors and their adjoints can always be chosen to be be orthogonal to each other; in addition, we'll choose the normalization

$$\mathbf{v}_l^\dagger(\theta) \cdot \mathbf{v}_k(\theta) = \delta_{lk}\,. \tag{11}$$

Letting

$$\delta\mathbf{r} = \sum_k a_k(t)\mathbf{v}_k(\theta_0)\,, \tag{12}$$

inserting this into Eq. (9), and using Eq. (10a), we arrive at a set of equations for the a_k,

$$\sum_k \dot{a}_k \mathbf{v}_k(\theta_0) = \sum_k a_k \lambda_k(\theta_0)\mathbf{v}_k(\theta_0) + \mathbf{S}(\theta_0). \tag{13}$$

Dotting both sides with $\mathbf{v}_l^\dagger(\theta_0)$, using Eq. (11), and then letting $l \to k$, we arrive at

$$\dot{a}_k = \lambda_k(\theta_0)a_k + \mathbf{v}_k^\dagger(\theta_0) \cdot \mathbf{S}(\theta_0)\,. \tag{14}$$

If we were linearizing around a fixed point, the next step would be to determine stability by finding the signs of the eigenvalues. However, because we have a line attractor (when h_i and V_{ij} are zero), there's a zero eigenvalues, which makes things a little harder. We have already said that the line attractor is stable. That means if the system gets a small push off the line attractor, it returns back. However, even under the unperturbed dynamics, Eq. (1), because of the zero eigenvalue it doesn't usually return to the same place.

Using $k = 0$ for the zero eigenvalue, the equation for a_0 is

$$\dot{a}_0 = \mathbf{v}_0^\dagger(\theta_0) \cdot \mathbf{S}(\theta_0). \tag{15}$$

This is just an integrator, and so the change in a_0 after a time dt is especially simple,

$$a_0(t + dt) = a_0(t) + dt\, \mathbf{v}_0^\dagger(\theta_0) \cdot \mathbf{S}(\theta_0). \tag{16}$$

Because a_0 integrates its input, under the unperturbed dynamics, for which there is no input (as $\mathbf{S}(\theta_0) = 0$), a_0 doesn't decay. This is in marked contrast to the other a_k, which do decay to zero when $\mathbf{S}(\theta_0) = 0$. Thus, when $\mathbf{S}(\theta_0) = 0$,

$$\mathbf{r}(t' \to \infty) = \mathbf{f}(\theta_0) + a_0(t)\mathbf{v}_0(\theta_0). \tag{17}$$

(Here $\mathbf{f}(\theta_0)$ is a vector whose i^{th} component is $f_i(\theta_0)$).

This suggests a natural way to choose θ_0: make sure that $a_0(t) = 0$ under the unperturbed dynamics. In general,

$$a_0(t) = \mathbf{v}_0(\theta_0) \cdot \big(\mathbf{r}(t) - \mathbf{f}(\theta_0)\big) \tag{18}$$

(see Eqs. (5) and (12), and use Eq. (11)); for $a_0(0)$ to be zero, we want to choose θ_0 so that

$$\mathbf{v}_0(\theta_0) \cdot \mathbf{r}(t) = \mathbf{v}_0(\theta_0) \cdot \mathbf{f}(\theta_0). \tag{19}$$

With this choice, $\mathbf{r}(t + dt)$ becomes

$$\mathbf{r}(t + dt) = \mathbf{f}(\theta_0) + dt\, \mathbf{v}_0(\theta_0)\mathbf{v}_0^\dagger(\theta_0) \cdot \mathbf{S}(\theta_0) + \sum_{k \neq 0} a_k(t + dt)\mathbf{v}_k(\theta_0) \tag{20}$$

where we used Eq. (16) and the fact that our choice of θ_0 ensured that $a_0(t) = 0$.

We now want to use the expression for $\mathbf{r}(t + dt)$ to determine the drift along the line attractor. To do that, we need to know what $\mathbf{v}_0$ is. It turns out that we can write down an explicit expression for it. Differentiating both sides of Eq. (2) gives us

$$f_i'(\theta) = \phi_i'(\theta) \sum_j W_{ij} f_j'(\theta) \tag{21}$$

where the prime on $f_i(\theta)$ denotes a derivative. This relationship implies that $f_i'(\theta)$ is an eigenvector of J_{ij} whose eigenvalue is zero (see Eq. (8a)). Thus, $v_{0i}(\theta_0) = f_i'(\theta_0)$; or, in vector notation, $\mathbf{v}_0(\theta_0) = \mathbf{f}'(\theta_0)$. Using this notation, $\mathbf{r}(t + dt)$ can be written (in the limit $dt \to 0$)

$$\begin{aligned} \mathbf{r}(t + dt) &= \mathbf{f}(\theta_0) + dt\, \mathbf{f}'(\theta_0)\mathbf{v}_0^\dagger(\theta_0) \cdot \mathbf{S}(\theta_0) + \sum_{k \neq 0} a_k(t + dt)\mathbf{v}_k(\theta_0) \\ &= \mathbf{f}\big(\theta_0 + dt\, \mathbf{v}_0^\dagger(\theta_0) \cdot \mathbf{S}(\theta_0)\big) + \sum_{k \neq 0} a_k(t + dt)\mathbf{v}_k(\theta_0). \end{aligned} \tag{22}$$

This gives us the following set of equations,

$$\mathbf{r}(t + dt) = \mathbf{f}(\theta) + \sum_{k \neq 0} a_k(t + dt)\mathbf{v}_k(\theta) \tag{23a}$$

$$\dot{\theta} = \mathbf{v}_0^\dagger \cdot \mathbf{S}(\theta) \tag{23b}$$

$$\theta(t) = \theta_0\,. \tag{23c}$$

The starting time, t, is arbitrary. Thus, the differential equation for θ holds for all time. In fact, Eq. (23b) is the important result; we don't care very much how the other a_k evolve.

If the right hand side of Eq. (23b) doesn't change sign with θ, then the trajectories will drift along the attractor forever; if the right hand side does change sign, with negative slope, then θ can be pulled to a fixed point. In either case, we don't have a line attractor: in the former we have a limit cycle (if θ is periodic); in the latter we have one or more fixed points.

So how can we control drift along the line attractor? The natural thing to do is provide appropriate external input. To investigate what we mean by "appropriate," we isolate the external input by letting $V_{ij} = 0$. Then, using Eq. (8b), we see that

$$\mathbf{v}_0^\dagger(\theta) \cdot \mathbf{S}(\theta) = \sum_i v_{0i}^\dagger(\theta)\phi_i'(\theta)h_i\,. \tag{24}$$

To make sense of this equation, we need to know what $v_{0i}^\dagger(\theta)$ looks like. It obeys the equation (dropping the θ-dependence for clarity)

$$\sum_j v_{0j}^\dagger(\theta)\phi_j'(\theta)W_{ji} = v_{0i}^\dagger(\theta)\,. \tag{25}$$

We can't make much progress in the general case, but we can if W_{ij} is symmetric, so that's what we'll use. Multiplying both sides of Eq. (25) by ϕ_i' we see that if W_{ij} is symmetric, Eq. (25) can be written

$$\sum_j \phi_i'(\theta)W_{ij}\left(v_{0j}^\dagger(\theta)\phi_j'(\theta)\right) = \left(v_{0i}^\dagger(\theta)\phi_i'(\theta)\right)\,. \tag{26}$$

Comparing this to Eq. (21), we see that

$$v_{0i}^\dagger(\theta) = \frac{1}{Z(\theta)}\frac{f_i'(\theta)}{\phi_i'(\theta)} \tag{27}$$

where $Z(\theta)$ is chosen to ensure that $\mathbf{v}_0^\dagger \cdot \mathbf{f}' = 1$,

$$Z(\theta) = \sum_i \frac{f_i'(\theta)^2}{\phi_i'(\theta)} \tag{28}$$

(see Eq. (11) and note that $\mathbf{v}_0 = \mathbf{f}'$),

Combining Eqs. (23b), (24) and (27), we see that θ evolves according to

$$\dot\theta = \frac{1}{Z(\theta)}\sum_i f_i'(\theta)\,h_i\,. \tag{29}$$

If we want to easily control the drift, we want the right hand side to be independent of θ, without h_i depending on θ (because we don't want to have to know θ to control motion along the line attractor). A natural choice is h constant: $h_i = h$. In that case, θ evolves according to

$$\dot\theta = h\,\frac{1}{Z(\theta)}\sum_i f_i'(\theta)\,. \tag{30}$$

This would be ideal: by modifying external input, h, we could induce drift along the line attractor at any rate we wanted. However, as we'll see shortly, for the "standard" line attractor at least, the sum over i vanishes, so this won't work.

Alternatively, we could use the recurrent connectivity, V_{ij}, to induce drift along the attractor. Letting $h_i = 0$,

$$\dot{\theta} = \sum_{ij} v^\dagger_{0i}(\theta)\phi'_i(\theta)V_{ij}f_j(\theta) \tag{31}$$

(see Eqs. (23b) and (8b)). Using Eq. (27), this becomes

$$\dot{\theta} = \frac{1}{Z(\theta)} \sum_{ij} f'_i(\theta)V_{ij}f_j(\theta)\,. \tag{32}$$

Here there is slightly more hope: letting, in a slight abuse of notation, $V_{ij} \to g(t)V_{ij}$, we have

$$\dot{\theta} = g(t)\,\frac{1}{Z(\theta)} \sum_{ij} f'_i(\theta)V_{ij}f_j(\theta)\,. \tag{33}$$

Now we just have to make the right hand side independent of θ, again with V_{ij} independent of θ. As we'll see in the next section, at least under the standard model, that's possible. However, while it's straightforward to encode $g(t)$ in neural activity (from, say, another network in the brain), the interaction is multiplicative, and so not all that easy to do with real neurons. An alternative approach is given in the homework.

The standard model for line attractors

As mentioned above, constructing a line attractor for an arbitrary weight matrix is difficult. However, it's relatively trivial for a translation invariant weight matrix: $W_{ij} = W_{i-j}$. In that case, Eq. (2) becomes

$$f(\theta_i - \theta) = \phi\left(\sum_j W_{i-j} f(\theta_j - \theta)\right) \tag{34}$$

where the θ_i are equally spaced, say with spacing $\Delta\theta$. It's easy to see that if this equation is satisfied for $\theta = \theta_0$, it's satisfied for $\theta = \theta_0 + k\Delta\theta$ with k an integer. For convenience, we'll assume it's satisfied for arbitrary (non-integer) k, making this a true line attractor. We'll also assume that θ is a periodic variable – which is kind of important; otherwise, things break down at the edges. And finally, we'll assume that W_{i-j} is symmetric: $W_{i-j} = W_{j-i}$. This is the standard model for periodic line attractors. We can now investigate our schemes above for inducing drift.

To simplify the analysis, we'll take the continuum limit. Basically, we take N to ∞, so the matrix multiplication in Eq. (2) turns into a convolution,

$$\sum_j W_{i-j} f(\theta_j - \theta) \to \sum_j W(\theta_i - \theta_j) f(\theta_j - \theta) \approx \frac{1}{\Delta\theta}\int d\alpha\, W(\theta_i - \alpha) f(\alpha - \theta)\,. \tag{35}$$

Defining the rescaled weights

$$w(\theta) = \frac{W(\theta)}{\Delta\theta}, \tag{36}$$

we have

$$\sum_j W_{i-j} f(\theta_j - \theta) \approx \int d\alpha\, w(\theta_i - \alpha) f(\alpha - \theta)\,. \tag{37}$$

Thus, after a small amount of algebra, Eq. (2) becomes

$$f(\theta) = \phi\left(\int d\alpha\, w(\theta - \alpha) f(\alpha)\right)\,. \tag{38}$$

We'll focus first on drift induced by h_i, with $h_i = h$ (Eq. (30)). For this case, the sum over i on the right hand side of Eq. (30) becomes

$$\sum_i f_i'(\theta) \to \frac{1}{\Delta\theta} \int d\alpha\, f'(\alpha - \theta) = \frac{1}{\Delta\theta}[f(\alpha_{\max} - \theta) - f(\alpha_{\min} - \theta)]. \tag{39}$$

Because θ is periodic, $\alpha_{\max} = \alpha_{\min}$, and the last term is zero. So if h_i is independent of i, the weight matrix is translation invariant and the connectivity is symmetric (this last one is needed to write down the adjoint eigenvalue), there's no easy way to use external input to control the position on the attractor.

Now we'll assume $h_i = 0$ and ask about the effect of V_{ij}. Letting

$$V_{ij} = \frac{v(\theta_i, \theta_j)}{\Delta\theta^2}\,, \tag{40}$$

Eq. (33) becomes

$$\dot{\theta} = g(t)\, \frac{1}{Z(\theta)} \int d\alpha d\beta\, f'(\alpha - \theta) v(\alpha, \beta) f(\beta - \theta) \tag{41}$$

where, using Eq. (7) for $\phi_i'(\theta)$, $Z(\theta)$ is given by

$$Z(\theta) \equiv \int d\alpha\, \frac{f'(\alpha - \theta)^2}{\phi'\left(\int d\beta w(\alpha - \beta) f(\beta - \theta)\right)} \tag{42}$$

Let $v(\alpha, \beta)$ be translation invariant, $v(\alpha, \beta) = v(\alpha - \beta)$, leading to

$$\dot{\theta} = g(t)\, \frac{1}{Z(\theta)} \int d\alpha d\beta f'(\alpha - \theta) v(\alpha - \beta) f(\beta - \theta)\,. \tag{43}$$

The integral is clearly independent of θ, as we can see by making the change of variables $\alpha \to \alpha + \theta$ and $\beta \to \beta + \theta$. Similarly the normalizer, $Z(\theta)$, is independent of θ, which we can see by making the same change of variables in Eq. (42). Thus, $\dot{\theta} \propto g(t)$. We also need to make sure that the integral over α and β isn't zero. We can do that by, for example, making

v an odd function of its argument. In that case, the integral over β is an approximation to a derivative, and so the integral over α is approximately the integral of the square of the derivative of f, which is positive.

The less standard model for line attractors

To perform this analysis, we needed W_{i-j} to be symmetric, so that we could write down an explicit expression for the adjoint eigenvector, $\mathbf{v}_0^\dagger$. What happens when W_{i-j} is not symmetric? The short answer is that I don't know. But I can speculate.

I have two beliefs in this case, neither of which I'm sure about. The first is that $Z(\theta)$ is still independent of θ. The second is that

$$\sum_i v_{0i}^\dagger(\theta)\phi_i'(\theta) \neq 0\,. \tag{44}$$

It would be nice to know whether either of these is true, but I haven't been able to figure that out. But let's say they are. We are then left with another question: is it possible to build a line attractor when W_{i-j} is not symmetric? Unfortunately, I don't know the answer to that one either. But at least I can outline the issues.

The approach is to choose the weight matrix, w, and the equilibrium function, f, and solve for the gain function, ϕ. If we can't find a gain function that satisfies Eq. (38) for any function f, than that weight matrix doesn't admit a line attractor. To do that, it helps to make the definition

$$F(\theta) \equiv \int d\alpha\, w(\theta-\alpha) f(\alpha). \tag{45}$$

Then, Eq. (38) can be written

$$\phi(F(\theta)) = f(\theta)\,. \tag{46}$$

Because we know the mapping from $f(\theta)$ to $F(\theta)$, we know ϕ evaluated at all points of interest – or at least all points of interest for constructing the line attractor. For other points we can choose ϕ arbitrarily.

Equation (46) must be satisfied for all θ, which isn't quite as trivial as it seems. That's because there are typically more than one value of θ that yield the same value of $F(\theta)$. For those values, it's $f(\theta)$ must also be the same. Put more succinctly, we have the condition

$$\text{if } F(\theta_1) = F(\theta_2), \text{ then } f(\theta_1) \text{ must equal } f(\theta_2). \tag{47}$$

If $f(\theta)$ is symmetric, $f(\theta) = f(-\theta)$, w is symmetric, $w(\theta) = w(-\theta)$, and both f and F are unimodal, then Eq. (47) is satisfied. That takes a little thinking, but a few well-drawn figures should convince you. The key observation is that if both f and w are symmetric, then F and f peak in the same place, and they're both symmetric around that peak. Add to that the unimodal constraint, and everything works out nicely.

Now we get to the question posed above: if $w(\theta)$ is not symmetric, is there some function f and nonlinearity ϕ such that Eq. (46) holds for all θ? Although I said in class that the answer is no, I'm now much less sure. What I showed is that if you pick an arbitrary function

f, even a unimodal one, it's very unlikely to satisfy Eq. (46). But that's not the question; we want to know if, for fixed non-symmetric $w(\theta)$, there's a ϕ and f for which Eq. (46) is satisfied for all θ. I'm actually leaning toward yes, but I have not been able to find an explicit construction. This will have to wait for the exam. ;)

Finally, I should point out that we did not have to assume that w is translation invariant; we could have performed all of the analysis in this section with an arbitrary weight matrix, of the form $w(\theta, \alpha)$. But that would make things even harder.