

---

# Supplementary Material for the paper "Neurons as Detectors of Coherent Sets in Sensory Dynamics"

---

Joshua L. Pughe-Sanford<sup>1,\*</sup> Xuehao Ding<sup>1,\*</sup> Jason J. Moore<sup>1,2,\*</sup> Anirvan M. Sengupta<sup>1,3</sup>  
Charles Epstein<sup>1</sup> Philip Greengard<sup>1</sup> Dmitri B. Chklovskii<sup>1,2</sup>

<sup>1</sup>Center for Computational Neuroscience, Flatiron Institute, Simons Foundation, New York, NY, USA

<sup>2</sup>Neuroscience Institute, NYU Langone Medical Center, New York, NY, USA

<sup>3</sup>Physics Department, Rutgers University, New Brunswick, NJ, USA

{jpughesanford, xding, cepstein, pgreengard, mitya}@flatironinstitute.org,  
jason.moore@nyulangone.org, anirvans.physics@gmail.com \* Equal contribution

## 1 Introduction to transfer operators

This is only a concise primer, for more information, see [1, 2].

### 1.1 Discrete case

We start from the discrete Markov chain for simplicity. Let  $\mathcal{X} = \{x_1, \dots, x_N\}$ . A one-step Markov transition is given by the kernel (matrix)

$$P(y | x), \quad \sum_y P(y | x) = 1.$$

Given an *initial* distribution  $\mu$  (row vector with  $\mu(x) \geq 0$ ,  $\sum_x \mu(x) = 1$ ), the *final* distribution after one step is

$$\nu(y) = \sum_x \mu(x) P(y | x).$$

For an observable  $f : \mathcal{X} \rightarrow \mathbb{R}$ , the stochastic Koopman operator is

$$(\mathcal{K}f)(x) = \sum_y P(y | x) f(y).$$

Equip the input and output spaces with the weighted inner products

$$\langle u, v \rangle_\mu = \sum_x u(x) v(x) \mu(x), \quad \langle u, v \rangle_\nu = \sum_y u(y) v(y) \nu(y).$$

The *adjoint* of  $\mathcal{K}$  w.r.t.  $(\mu, \nu)$  is the operator  $\mathcal{K}^\dagger : L^2(\mu) \rightarrow L^2(\nu)$  defined by

$$\langle \mathcal{K}f, g \rangle_\mu = \langle f, \mathcal{K}^\dagger g \rangle_\nu \quad \text{for all } f, g.$$

A direct computation yields, for any  $y$  with  $\nu(y) > 0$ ,

$$(\mathcal{K}^\dagger g)(y) = \frac{1}{\nu(y)} \sum_x \mu(x) P(y | x) g(x). \quad (1)$$

In matrix form, with  $D_\mu = \text{diag}(\mu)$  and  $D_\nu = \text{diag}(\nu)$ ,

$$\mathcal{K}^\dagger = D_\nu^{-1} P^\top D_\mu.$$

The reverse-time kernel proposed by [3] is given by

$$\tilde{P}(x | y) = \frac{\mu(x) P(y | x)}{\nu(y)}.$$

Then (1) can be rewritten as

$$(\mathcal{K}^\dagger g)(y) = \sum_x \tilde{P}(x | y) g(x),$$

i.e.,  $\mathcal{K}^\dagger$  is precisely the *Koopman operator for the reverse-time Markov step*  $y \mapsto x \sim \tilde{P}(\cdot | y)$  associated with pushing  $\mu$  forward to  $\nu = \mu P$ .

**Special cases.** If  $\mu = \nu = \pi$  is stationary and the chain is reversible ( $\pi(x)P(y | x) = \pi(y)P(x | y)$ ), then  $\tilde{P}(x | y) = P(x | y)$  and  $\mathcal{K}^\dagger = \mathcal{K}$  (self-adjointness in  $L^2(\pi)$ ).

## 1.2 Continuous case

Let  $(\mathcal{K}_t)_{t \geq 0}$  be the Koopman semigroup with transition density  $p_t(y | x)$ :

$$(\mathcal{K}_t f)(x) = \mathbb{E}[f(X_t) | X_0 = x] = \int f(y) p_t(y | x) dy,$$

$$\mathcal{K}_t \mathcal{K}_s = \mathcal{K}_{t+s}.$$

Given an initial distribution  $\mu$  (measure or density), the pushed-forward distribution at time  $t$  is

$$\nu_t(dy) = \int \mu(dx) p_t(y | x) dy.$$

With  $\langle u, v \rangle_\mu = \int u(x)v(x) \mu(dx)$  and  $\langle u, v \rangle_{\nu_t} = \int u(y)v(y) \nu_t(dy)$ , the  $(\mu, \nu_t)$ -weighted adjoint  $\mathcal{K}_t^\dagger : L^2(\mu) \rightarrow L^2(\nu_t)$  satisfies

$$\langle \mathcal{K}_t f, g \rangle_\mu = \langle f, \mathcal{K}_t^\dagger g \rangle_{\nu_t},$$

and is given by the reverse-time Koopman operator

$$\mathcal{K}_t^\dagger g(y) = \int g(x) \tilde{p}_t(x | y) dx, \quad \tilde{p}_t(x | y) = \frac{\mu(x) p_t(y | x)}{\nu_t(y)}. \quad (2)$$

Thus the adjoint w.r.t.  $(\mu, \nu_t)$  is the Koopman operator for the time-reversed process at horizon  $t$  determined by  $\mu$ .

Consider an SDE  $dX_t = a(X_t) dt + \sqrt{D} dW_t$ . The generator of the Koopman operator can be calculated using Itô's formula as

$$\begin{aligned} & \lim_{t \rightarrow 0} \frac{\mathcal{K}_t f(x) - f(x)}{t} \\ &= \mathbb{E} \left[ \lim_{t \rightarrow 0} \frac{f(X_t) - f(x)}{t} \middle| X_0 = x \right] \\ &= a(x) \cdot \nabla f(x) + \frac{1}{2} \nabla \cdot (D \nabla f(x)) \\ &= \mathcal{L}^\dagger f(x). \end{aligned} \quad (3)$$

Thus, we obtain the equality  $\mathcal{K}_t = e^{\mathcal{L}^\dagger t}$ .

## 2 The invariant measure

Here we prove that the distribution given by Eq. (14, main text) is invariant under the action of the forward Kolmogorov operator, Eq. (11, main text):

$$\begin{aligned}
\mathcal{L}\rho_0(\mathbf{x}) &= -\nabla \cdot [A\mathbf{x}\rho_0(\mathbf{x})] + \frac{1}{2}\nabla \cdot [D\nabla\rho_0(\mathbf{x})] \\
&= [-\text{Tr}(A) + \mathbf{x}^\top A^\top \Sigma^{-1}\mathbf{x} - \frac{1}{2}\text{Tr}(D\Sigma^{-1}) + \frac{1}{2}\mathbf{x}^\top \Sigma^{-1}D\Sigma^{-1}\mathbf{x}]\rho_0(\mathbf{x}) \\
&= [-\text{Tr}(A) - \frac{1}{2}\text{Tr}(D\Sigma^{-1})]\rho_0(\mathbf{x}) \\
&= -\frac{1}{2}\text{Tr}(A - \Sigma A^\top \Sigma^{-1})\rho_0(\mathbf{x}) \\
&= 0,
\end{aligned} \tag{4}$$

where the Lyapunov equation, Eq. (15, main text), was used to obtain the 3rd and the 4th line.

## 3 Singular Spectrum of the SKO for OU processes

In order to find the singular functions of the SKO, we search for the eigenfunctions of the Forward-backward and the Backward-Forward operators:

$$\mathcal{F}_\tau = e^{\mathcal{L}^\dagger \tau} \text{diag}(\rho_0)^{-1} e^{\mathcal{L} \tau} \text{diag}(\rho_0) \quad \mathcal{B}_\tau = \text{diag}(\rho_0)^{-1} e^{\mathcal{L} \tau} \text{diag}(\rho_0) e^{\mathcal{L}^\dagger \tau} \tag{5}$$

where  $\text{diag}(\rho_0)^{-1} e^{\mathcal{L} \tau} \text{diag}(\rho_0)$  is termed the reweighted Perron-Frobenius operator, which shares the domain with the SKO.

### 3.1 Linear eigenfunctions

We start by considering candidate eigenfunctions that are linear w.r.t. the input state:

$$q_\theta(\mathbf{x}) = \theta^\top \mathbf{x}, \tag{6}$$

where  $\theta \in \mathbb{R}^n$ . The actions of  $\text{diag}(\rho_0)^{-1} \mathcal{L} \text{diag}(\rho_0)$  and  $\mathcal{L}^\dagger$  on  $q_\theta(\mathbf{x})$  are given by

$$\begin{aligned}
&\text{diag}(\rho_0)^{-1} \mathcal{L} \text{diag}(\rho_0) q_\theta(\mathbf{x}) \\
&= \rho_0(\mathbf{x})^{-1} [-\nabla \cdot (A\mathbf{x}\rho_0(\mathbf{x})\theta^\top \mathbf{x}) + \frac{1}{2} D^{\alpha\beta} \partial_\alpha \partial_\beta (\rho_0(\mathbf{x})\theta^\top \mathbf{x})] \\
&= -\theta^\top A\mathbf{x} - \mathbf{x}^\top \Sigma^{-1} D \theta + \rho_0(\mathbf{x})^{-1} \theta^\top \mathbf{x} \mathcal{L} \rho_0(\mathbf{x}) \\
&= (\Sigma^{-1} A \Sigma \theta)^\top \mathbf{x},
\end{aligned} \tag{7}$$

$$\begin{aligned}
&\mathcal{L}^\dagger q_\theta(\mathbf{x}) \\
&= (A\mathbf{x})^\top \nabla (\theta^\top \mathbf{x}) \\
&= (A^\top \theta)^\top \mathbf{x},
\end{aligned} \tag{8}$$

where the Lyapunov equation, Eq. (15, main text), and invariant density annihilation, Eq.(4), were used.

Then the actions of  $\text{diag}(\rho_0)^{-1} e^{\mathcal{L} \tau} \text{diag}(\rho_0)$  and  $e^{\mathcal{L}^\dagger \tau}$  on  $q_\theta(\mathbf{x})$  are given by

$$\begin{aligned}
&\text{diag}(\rho_0)^{-1} e^{\mathcal{L} \tau} \text{diag}(\rho_0) q_\theta(\mathbf{x}) \\
&= \text{diag}(\rho_0)^{-1} \sum_{i=0}^{\infty} \frac{\mathcal{L}^i \tau^i}{i!} \text{diag}(\rho_0) q_\theta(\mathbf{x}) \\
&= \left( \sum_{i=0}^{\infty} \frac{\Sigma^{-1} A^i \tau^i \Sigma}{i!} \theta \right)^\top \mathbf{x} \\
&= (\Sigma^{-1} e^{A\tau} \Sigma \theta)^\top \mathbf{x},
\end{aligned} \tag{9}$$

$$\begin{aligned}
& e^{\mathcal{L}^\dagger \tau} q_\theta(\mathbf{x}) \\
&= \sum_{i=0}^{\infty} \frac{\mathcal{L}^{\dagger i} \tau^i}{i!} q_\theta(\mathbf{x}) \\
&= \left( \sum_{i=0}^{\infty} \frac{(A^\top)^i \tau^i}{i!} \theta \right)^\top \mathbf{x} \\
&= (e^{A^\top \tau} \theta)^\top \mathbf{x}.
\end{aligned} \tag{10}$$

Combining Eqs.(9)(10), denoting the matrices  $\exp(A^\top \tau) \Sigma^{-1} \exp(A \tau) \Sigma$  and  $\Sigma^{-1} \exp(A \tau) \Sigma \exp(A^\top \tau)$  by  $P$  and  $Q$ , respectively, we have

$$\mathcal{F}_\tau q_\theta(\mathbf{x}) = q_{P\theta}(\mathbf{x}), \quad \mathcal{B}_\tau q_\theta(\mathbf{x}) = q_{Q\theta}(\mathbf{x}). \tag{11}$$

Rewriting Eq. (18, main text) as

$$P\mathbf{v}_i = \lambda_i^2 \mathbf{v}_i, \quad Q\mathbf{u}_i = \lambda_i^2 \mathbf{u}_i, \tag{12}$$

we obtain a set of linear eigenfunctions of the forward-backward and the backward-forward operators:

$$\mathcal{F}_\tau q_{v_i}(\mathbf{x}) = \lambda_i^2 q_{v_i}(\mathbf{x}), \quad \mathcal{B}_\tau q_{u_i}(\mathbf{x}) = \lambda_i^2 q_{u_i}(\mathbf{x}). \tag{13}$$

So long as each set of eigenvectors are linearly independent, this method obtains all linear eigenfunctions  $\mathcal{F}_\tau$  and  $\mathcal{B}_\tau$ . This is because the dimensionality of the subspace spanned by linear functions is equal to  $n$ , which equals the number of eigenfunctions found.

We note that our calculations apply regardless of the stability of the dynamics, i.e. whether the real parts of the eigenvalues of  $\mathbf{A}$  are positive or negative. For stable  $\mathbf{A}$ , all  $\{\lambda_i^2\}$  are less than one. The sign of the eigenfunction with the largest  $\lambda_i^2$  partitions the state space into the minimally mixing coherent sets relative to the invariant density. For unstable dynamics, an invariant density in the form of Eq. 14 of main text persists (albeit not normalizable) and all  $\{\lambda_i^2\}$  are greater than one. In this setting, the sign of the eigenfunction with the smallest  $\lambda_i^2$  partitions the state space into the minimally mixing coherent sets relative to the invariant density. Intuitively, one can view the unstable forward dynamics as a stable dynamics backwards in time of the dual variables [3, 4, 5].

### 3.2 Quadratic eigenfunctions

Here, we consider candidate eigenfunctions that are up to the second order in the input state:

$$q_{M;c}(\mathbf{x}) = \mathbf{x}^\top M \mathbf{x} + c, \tag{14}$$

where  $M$  is an  $n \times n$  symmetric matrix,  $c$  is a scalar. It will soon become clear why we need an extra constant scalar rather than a linear term in the function. Since  $\mathcal{L}c = \mathcal{L}^\dagger c = 0$ ,  $q_M(\mathbf{x})$  will be used in the derivation below for brevity.

We first determine the actions of  $\text{diag}(\rho_0)^{-1} \mathcal{L} \text{diag}(\rho_0)$  and  $\mathcal{L}^\dagger$  on  $q_M(\mathbf{x})$ .

$$\begin{aligned}
& \text{diag}(\rho_0)^{-1} \mathcal{L} \text{diag}(\rho_0) q_M(\mathbf{x}) \\
&= \rho_0(\mathbf{x})^{-1} [-\nabla \cdot (A \mathbf{x} \rho_0(\mathbf{x}) \mathbf{x}^\top M \mathbf{x}) + \frac{1}{2} D^{\alpha\beta} \partial_\alpha \partial_\beta (\rho_0(\mathbf{x}) \mathbf{x}^\top M \mathbf{x})] \\
&= -2\mathbf{x}^\top M A \mathbf{x} - 2\mathbf{x}^\top \Sigma^{-1} D M \mathbf{x} + \text{Tr}(DM) \\
&= 2\mathbf{x}^\top \Sigma^{-1} A \Sigma M \mathbf{x} - 2 \text{Tr}(A \Sigma M) \\
&= \mathbf{x}^\top (\Sigma^{-1} A \Sigma M + M \Sigma A^\top \Sigma^{-1}) \mathbf{x} - 2 \text{Tr}(A \Sigma M),
\end{aligned} \tag{15}$$

$$\begin{aligned}
& \mathcal{L}^\dagger q_M(\mathbf{x}) \\
&= (A \mathbf{x})^\top \nabla (\mathbf{x}^\top M \mathbf{x}) + \frac{1}{2} D^{\alpha\beta} \partial_\alpha \partial_\beta (\mathbf{x}^\top M \mathbf{x}) \\
&= 2\mathbf{x}^\top A^\top M \mathbf{x} - 2 \text{Tr}(A \Sigma M) \\
&= \mathbf{x}^\top (A^\top M + M A) \mathbf{x} - 2 \text{Tr}(A \Sigma M),
\end{aligned} \tag{16}$$

where the Lyapunov equation and the fact  $\mathcal{L}\rho_0 = 0$  have been used.

We noticed that

$$\begin{aligned}
(\mathcal{L}^\dagger)^2 q_M(\mathbf{x}) &= \mathbf{x}^\top ((A^\top)^2 M + 2A^\top M A + M A^2) \mathbf{x} + \text{const.} \\
(\mathcal{L}^\dagger)^3 q_M(\mathbf{x}) &= \mathbf{x}^\top ((A^\top)^3 M + 3(A^\top)^2 M A + 3A^\top M A^2 + M A^3) \mathbf{x} + \text{const.} \\
&\dots \\
(\mathcal{L}^\dagger)^m q_M(\mathbf{x}) &= \mathbf{x}^\top \left( \sum_{i=0}^m C_m^i (A^\top)^{m-i} M A^i \right) \mathbf{x} + \text{const.}
\end{aligned} \tag{17}$$

Thus, the action of  $e^{\mathcal{L}^\dagger \tau}$  on  $q_M(\mathbf{x})$  is given by

$$\begin{aligned}
e^{\mathcal{L}^\dagger \tau} q_M(\mathbf{x}) &= \mathbf{x}^\top \left( \sum_{m=0}^{\infty} \frac{\tau^m}{m!} \sum_{i=0}^m C_m^i (A^\top)^{m-i} M A^i \right) \mathbf{x} + \text{const.} \\
&= \mathbf{x}^\top \left( \sum_{i=0}^{\infty} \sum_{m=i}^{\infty} \frac{(A^\top)^{m-i} \tau^{m-i}}{(m-i)!} M \frac{A^i \tau^i}{i!} \right) \mathbf{x} + \text{const.} \\
&= \mathbf{x}^\top (e^{A^\top \tau} M e^{A \tau}) \mathbf{x} + \text{const.}
\end{aligned} \tag{18}$$

Similarly, the action of  $\text{diag}(\rho_0)^{-1} e^{\mathcal{L} \tau} \text{diag}(\rho_0)$  on  $q_M(\mathbf{x})$  is given by

$$\text{diag}(\rho_0)^{-1} \mathcal{L}^m \text{diag}(\rho_0) q_M(\mathbf{x}) = \mathbf{x}^\top \left( \sum_{i=0}^m C_m^i \Sigma^{-1} A^{m-i} \Sigma M \Sigma (A^\top)^i \Sigma^{-1} \right) \mathbf{x} + \text{const.} \tag{19}$$

$$\text{diag}(\rho_0)^{-1} e^{\mathcal{L} \tau} \text{diag}(\rho_0) q_M(\mathbf{x}) = \mathbf{x}^\top (\Sigma^{-1} e^{A \tau} \Sigma M \Sigma e^{A^\top \tau} \Sigma^{-1}) \mathbf{x} + \text{const.} \tag{20}$$

Finally we have the actions of  $\mathcal{F}_\tau$  and  $\mathcal{B}_\tau$  on  $q_M(\mathbf{x})$ :

$$\mathcal{F}_\tau q_M(\mathbf{x}) = \mathbf{x}^\top (P M P^\top) \mathbf{x} + \text{const.} \tag{21}$$

$$\mathcal{B}_\tau q_M(\mathbf{x}) = \mathbf{x}^\top (Q M Q^\top) \mathbf{x} + \text{const.} \tag{22}$$

With careful choice of the constant term, we obtain a set of quadratic eigenfunctions

$$\mathcal{F}_\tau q_{v_i v_i^\top}(\mathbf{x}) = \lambda_i^4 q_{v_i v_i^\top}(\mathbf{x}), \tag{23}$$

$$\mathcal{F}_\tau q_{v_i v_j^\top + v_j v_i^\top}(\mathbf{x}) = \lambda_i^2 \lambda_j^2 q_{v_i v_j^\top + v_j v_i^\top}(\mathbf{x}),$$

$$\mathcal{B}_\tau q_{u_i u_i^\top}(\mathbf{x}) = \lambda_i^4 q_{u_i u_i^\top}(\mathbf{x}), \tag{24}$$

$$\mathcal{B}_\tau q_{u_i u_j^\top + u_j u_i^\top}(\mathbf{x}) = \lambda_i^2 \lambda_j^2 q_{u_i u_j^\top + u_j u_i^\top}(\mathbf{x}).$$

Since the number of eigenfunctions  $\binom{n(n+1)}{2}$  is equal to the dimensionality of the subspace spanned by  $n \times n$  symmetric matrices, we obtained all quadratic eigenfunctions. For attractive dynamics,  $0 < \lambda_i^2 < 1$ , so their eigenvalues are less than the greatest eigenvalue of linear functions. For repulsive dynamics,  $\lambda_i^2 > 1$ , so their eigenvalues are greater than the smallest eigenvalue of linear functions.

### 3.3 General situation

On  $\mathbb{R}^n$  consider a symmetric  $k$ -linear form  $M_k : (\mathbb{R}^n)^{\otimes k} \rightarrow \mathbb{R}$  such that  $M_k(v_1 \otimes v_2 \otimes \dots \otimes v_k) = M_k(v_{\sigma(1)} \otimes v_{\sigma(2)} \otimes \dots \otimes v_{\sigma(k)})$  for any permutation  $\sigma \in S_k$ . Its matrix representation is given by  $M_k^{I_1 I_2 \dots I_k} = M_k(e_{I_1} \otimes e_{I_2} \otimes \dots \otimes e_{I_k})$  where  $\{e_i : i = 1, 2, \dots, n\}$  is the standard orthonormal basis of  $\mathbb{R}^n$ . Then we can consider  $k$ -th order tentative eigenfunctions of the form:

$$q_M^k(\mathbf{x}) = M_k(\mathbf{x}^{\otimes k}) + O(k-2), \tag{25}$$

where  $O(k-2)$  denotes a (multivariate) polynomial up to the  $(k-2)$ -th order. Similar to the quadratic case, the eigenvalue is solely determined by the leading term. Thus, it is sufficient to analyze the actions of operators on  $M_k$ .

The actions of  $\text{diag}(\rho_0)^{-1} \mathcal{L} \text{diag}(\rho_0)$  and  $\mathcal{L}^\dagger$  on  $M_k(\mathbf{x}^{\otimes k})$  are given by

$$\begin{aligned}
& \text{diag}(\rho_0)^{-1} \mathcal{L} \text{diag}(\rho_0) M_k(\mathbf{x}^{\otimes k}) \\
&= \rho_0(\mathbf{x})^{-1} [-\nabla \cdot (A \mathbf{x} \rho_0(\mathbf{x}) M_k(\mathbf{x}^{\otimes k})) + \frac{1}{2} \nabla \cdot (D \nabla \rho_0(\mathbf{x}) M_k(\mathbf{x}^{\otimes k}))] \\
&= -\langle A \mathbf{x}, \nabla M_k(\mathbf{x}^{\otimes k}) \rangle + \langle \nabla \log \rho_0(\mathbf{x}), D \nabla M_k(\mathbf{x}^{\otimes k}) \rangle + \frac{1}{2} \nabla \cdot (D \nabla M_k(\mathbf{x}^{\otimes k})) \\
&= \langle -(A + D \Sigma^{-1}) \mathbf{x}, \nabla M_k(\mathbf{x}^{\otimes k}) \rangle + O(k-2) \\
&= \langle \Sigma A^\top \Sigma^{-1} \mathbf{x}, \nabla M_k(\mathbf{x}^{\otimes k}) \rangle + O(k-2) \\
&= \sum_{i=1}^k M_k \times_i (\Sigma^{-1} A \Sigma)(\mathbf{x}^{\otimes k}) + O(k-2),
\end{aligned} \tag{26}$$

$$\begin{aligned}
& \mathcal{L}^\dagger M_k(\mathbf{x}^{\otimes k}) \\
&= \langle A \mathbf{x}, \nabla M_k(\mathbf{x}^{\otimes k}) \rangle + \frac{1}{2} \nabla \cdot (D \nabla M_k(\mathbf{x}^{\otimes k})) \\
&= \sum_{i=1}^k M_k \times_i A^\top(\mathbf{x}^{\otimes k}) + O(k-2),
\end{aligned} \tag{27}$$

where the Lyapunov equation and the fact  $\mathcal{L} \rho_0 = 0$  have been used.  $\times_i$  denotes tensor mode product:  $(T \times_n A)_{i_1, \dots, i_{n-1}, j, i_{n+1}, \dots, i_k} \equiv \sum_{i_n} A_{j i_n} T_{i_1, \dots, i_{n-1}, i_n, i_{n+1}, \dots, i_k}$  satisfying the following algebra.

$$\begin{aligned}
T \times_i A \times_i B &= T \times_i (BA), \\
T \times_i A \times_j B &= T \times_j B \times_i A \quad i \neq j.
\end{aligned} \tag{28}$$

Then we can determine the actions of finite-time operators on  $M_k(\mathbf{x}^{\otimes k})$ .

$$(\mathcal{L}^\dagger)^m M_k(\mathbf{x}^{\otimes k}) = \sum_{i_m=1}^k \dots \sum_{i_2=1}^k \sum_{i_1=1}^k M_k \times_{i_1} A^\top \times_{i_2} A^\top \dots \times_{i_m} A^\top(\mathbf{x}^{\otimes k}) + O(k-2), \tag{29}$$

$$\begin{aligned}
& e^{\mathcal{L}^\dagger \tau} M_k(\mathbf{x}^{\otimes k}) \\
&= \sum_{m=0}^{\infty} \frac{\tau^m (\mathcal{L}^\dagger)^m}{m!} M_k(\mathbf{x}^{\otimes k}) \\
&= \sum_{\{l_1, l_2, \dots, l_{k-1}\}=0}^{\infty} \sum_{m=\sum_{i=1}^{k-1} l_i}^{\infty} \frac{\tau^m}{m!} \binom{m}{l_1, l_2, \dots, l_{k-1}, m - \sum_{i=1}^{k-1} l_i} \\
& \quad M_k \times_1 (A^\top)^{l_1} \times_2 (A^\top)^{l_2} \dots \times_{k-1} (A^\top)^{l_{k-1}} \times_k (A^\top)^{m - \sum_{i=1}^{k-1} l_i}(\mathbf{x}^{\otimes k}) + O(k-2) \\
&= M_k \times_1 e^{A^\top \tau} \times_2 e^{A^\top \tau} \dots \times_k e^{A^\top \tau}(\mathbf{x}^{\otimes k}) + O(k-2).
\end{aligned} \tag{30}$$

Similarly,

$$\begin{aligned}
& \text{diag}(\rho_0)^{-1} e^{\mathcal{L} \tau} \text{diag}(\rho_0) M_k(\mathbf{x}^{\otimes k}) \\
&= M_k \times_1 (\Sigma^{-1} e^{A \tau} \Sigma) \times_2 (\Sigma^{-1} e^{A \tau} \Sigma) \dots \times_k (\Sigma^{-1} e^{A \tau} \Sigma)(\mathbf{x}^{\otimes k}) + O(k-2).
\end{aligned} \tag{31}$$

Then the actions of  $\mathcal{F}_\tau$  and  $\mathcal{B}_\tau$  on  $M_k(\mathbf{x}^{\otimes k})$  follow as

$$\mathcal{F}_\tau M_k(\mathbf{x}^{\otimes k}) = M_k \times_1 P \times_2 P \dots \times_k P(\mathbf{x}^{\otimes k}) + O(k-2), \tag{32}$$

$$\mathcal{B}_\tau M_k(\mathbf{x}^{\otimes k}) = M_k \times_1 Q \times_2 Q \dots \times_k Q(\mathbf{x}^{\otimes k}) + O(k-2). \tag{33}$$

We therefore obtain a set of  $k$ -th-order eigenfunctions of  $\mathcal{F}_\tau$  and  $\mathcal{B}_\tau$ :

$$\mathcal{F}_\tau q_{\sum_{\sigma \in S_k} v_{I_{\sigma(1)}} \otimes v_{I_{\sigma(2)}} \otimes \dots \otimes v_{I_{\sigma(k)}}}^k(\mathbf{x}^{\otimes k}) = \left( \prod_{i=1}^k \lambda_{I_i}^2 \right) q_{\sum_{\sigma \in S_k} v_{I_{\sigma(1)}} \otimes v_{I_{\sigma(2)}} \otimes \dots \otimes v_{I_{\sigma(k)}}}^k(\mathbf{x}^{\otimes k}), \tag{34}$$

$$\mathcal{B}_\tau q_{\sum_{\sigma \in S_k} u_{I_{\sigma(1)}} \otimes u_{I_{\sigma(2)}} \otimes \dots \otimes u_{I_{\sigma(k)}}}^k(\mathbf{x}^{\otimes k}) = \left( \prod_{i=1}^k \lambda_{I_i}^2 \right) q_{\sum_{\sigma \in S_k} u_{I_{\sigma(1)}} \otimes u_{I_{\sigma(2)}} \otimes \dots \otimes u_{I_{\sigma(k)}}}^k(\mathbf{x}^{\otimes k}). \tag{35}$$

Again, these eigenvalues are less than the greatest eigenvalue of linear eigenfunctions for attractive dynamics and greater than the smallest eigenvalue of linear eigenfunctions for repulsive dynamics.

## 4 Orthogonality analysis of singular vectors for OU processes

This section proves that asymptotically the top predictive mode converges to the top left eigenvector and the bottom retrospective mode converges to the bottom left eigenvector. This helps us understand the 2d saddle-point example in Fig. 1 and provides analytical support for the orthogonality analysis of data in Figs. 2 and 3.

### 4.1 Top eigenvector

We start from the definitions of eigenvectors of transfer operators:

$$e^{A^\top \tau} \Sigma^{-1} e^{A\tau} \Sigma \mathbf{v}_i = \lambda_i^2 \mathbf{v}_i, \quad \Sigma^{-1} e^{A\tau} \Sigma e^{A^\top \tau} \mathbf{u}_i = \lambda_i^2 \mathbf{u}_i. \quad (36)$$

Notice that

$$\begin{aligned} & (e^{A^\top \tau} \Sigma^{-1} e^{A\tau} \Sigma)^\top \Sigma \mathbf{v}_i \\ &= \Sigma e^{A^\top \tau} \Sigma^{-1} e^{A\tau} \Sigma \mathbf{v}_i \\ &= \lambda_i^2 \Sigma \mathbf{v}_i. \end{aligned} \quad (37)$$

Thus,  $\{\mathbf{v}_i, \Sigma \mathbf{v}_i\}$  is a pair of right and left eigenvectors of  $e^{A^\top \tau} \Sigma^{-1} e^{A\tau} \Sigma$ . Similarly,  $\{\mathbf{u}_i, \Sigma \mathbf{u}_i\}$  is a pair of right and left eigenvectors of  $\Sigma^{-1} e^{A\tau} \Sigma e^{A^\top \tau}$ . Therefore,  $\mathbf{u}_i^\top \Sigma \mathbf{u}_j = \mathbf{v}_i^\top \Sigma \mathbf{v}_j = 0$  for  $i \neq j$ .

Also notice that

$$\begin{aligned} & e^{A^\top \tau} (\Sigma^{-1} e^{A\tau} \Sigma e^{A^\top \tau} \mathbf{u}_i) \\ &= (e^{A^\top \tau} \Sigma^{-1} e^{A\tau} \Sigma) e^{A^\top \tau} \mathbf{u}_i \\ &= \lambda_i^2 e^{A^\top \tau} \mathbf{u}_i. \end{aligned} \quad (38)$$

Thus,  $\mathbf{v}_i \propto e^{A^\top \tau} \mathbf{u}_i$ . Under proper normalization, these results can be summarized by

$$V^\top \Sigma V = Id, \quad U^\top \Sigma U = Id, \quad e^{A^\top \tau} U = V \Lambda, \quad (39)$$

where  $U[:, i] := \mathbf{u}_i$ ,  $V[:, i] := \mathbf{v}_i$ ,  $\Lambda_{ij} := \delta_{ij} \lambda_i$ . Thus, we have

$$e^{A^\top \tau} = V \Lambda U^{-1} = V \Lambda U^\top \Sigma, \quad (40)$$

$$e^{A\tau} = \Sigma U \Lambda V^\top = \sum_i \lambda_i \Sigma \mathbf{u}_i \mathbf{v}_i^\top. \quad (41)$$

On the other hand, we have the eigen-decomposition

$$e^{A\tau} = \sum_i e^{a_i \tau} \mathbf{r}_i \mathbf{l}_i^\top, \quad (42)$$

where  $a_i$  is the eigenvalue of  $A$ ,  $\mathbf{r}_i$  and  $\mathbf{l}_i$  are right and left eigenvectors of  $A$ , respectively. We consider the asymptotic limit as  $\tau \rightarrow \infty$ , where the leading rank-1 term of  $e^{A\tau}$  dominates. Comparing the leading terms of Eq. (41) and Eq. (42), we have

$$\mathbf{v}_1 \propto \mathbf{l}_1, \quad \Sigma \mathbf{u}_1 \propto \mathbf{r}_1. \quad (43)$$

Thus, the top predictive mode is orthogonal to every eigen-direction except the most repulsive direction. In particular,  $\mathbf{v}_1$  is orthogonal to the stable manifold.

### 4.2 Bottom eigenvector

Notice that under the transformation  $A \rightarrow -A^\top$ , we have

$$\begin{aligned} e^{A\tau} &\rightarrow [e^{A\tau}]^{-\top} \\ &= [\Sigma U \Lambda V^\top]^{-\top} \\ &= U \Lambda^{-1} V^\top \Sigma \\ &= \sum_i \lambda_i^{-1} \mathbf{u}_i (\Sigma \mathbf{v}_i)^\top. \end{aligned} \quad (44)$$

And the eigen-decomposition becomes

$$e^{-A^\top \tau} = \sum_i e^{-a_i \tau} \mathbf{l}_i \mathbf{r}_i^\top, \quad (45)$$

Similar to the first section, by matching the leading rank-1 term under the  $\tau \rightarrow \infty$  limit, we have

$$\Sigma \mathbf{v}_n \propto \mathbf{r}_n, \quad \mathbf{u}_n \propto \mathbf{l}_n. \quad (46)$$

Thus, the bottom retrospective mode is orthogonal to every eigen-direction except the most attractive direction. In particular,  $\mathbf{u}_n$  is orthogonal to the unstable manifold.

## 5 Galerkin approximation of transfer operators [2]

Real neurons are exposed to only a finite number of channels or synaptic inputs. Mathematically, this means that it can only have access to a finite collection of observables, which we write as the Galerkin projection, referring to the projection of the SKO onto the linear span of these accessible measurement functions. Let

$$\mathcal{F} = \{f(\mathbf{x}) \mid f(\mathbf{x}) = \sum_{i=1}^d c_i \phi_i(\mathbf{x}) = \vec{c} \cdot \vec{\phi}(\mathbf{x})\}, \quad (47)$$

denote the linear span of a finite set of basis functions  $\{\phi_i\}_{i=1}^d$ . Since each  $f \in \mathcal{F}$  is uniquely determined by the coefficient vector  $\vec{c} \in \mathbb{R}^d$ , the space  $\mathcal{F}$  is naturally isomorphic to  $\mathbb{R}^d$ . Given some inner product, any linear operator  $\mathcal{A} : \mathcal{F} \rightarrow \mathcal{F}$  admits a matrix representation  $A : \mathbb{R}^d \rightarrow \mathbb{R}^d$  satisfying

$$[\mathcal{A} \vec{c} \cdot \vec{\phi}](\mathbf{x}) = (A\vec{c}) \cdot \vec{\phi}(\mathbf{x}), \quad (48)$$

where  $A = (\Sigma)^{-1} \Sigma_{\mathcal{A}}$  and

$$\Sigma_{ij} = \langle \phi_i, \phi_j \rangle, \quad [\Sigma_{\mathcal{A}}]_{ij} = \langle \phi_i, \mathcal{A} \phi_j \rangle. \quad (49)$$

For functions  $f = \vec{c}_f \cdot \vec{\phi}$  and  $g = \vec{c}_g \cdot \vec{\phi}$  in  $\mathcal{F}$ , the inner product reduces to

$$\langle f, g \rangle = \vec{c}_f \cdot \Sigma \vec{c}_g. \quad (50)$$

This motivates the Galerkin projection of the SKO, which is a linear operator describing the application of  $\mathcal{K}_\tau$ , followed by projection onto  $\mathcal{F}$ . This operator maps  $\mathcal{F}$  to itself, and admits a matrix representation, [2, 6]

$$K_\tau = (\Sigma)^{-1} \Sigma_{\mathcal{K}_\tau}, \quad (51)$$

where  $\phi_i$  are assumed to be linearly independent such that  $\Sigma$  is invertible. To specify  $K_\tau$  entirely, one must define the inner product.

For a statistically stationary process with invariant density  $\rho(\mathbf{x})$ , a natural inner product is defined as

$$\langle f, g \rangle = \int_{\mathcal{X}} f(\mathbf{x}) g(\mathbf{x}) \rho(\mathbf{x}) d\mathbf{x} = \mathbb{E}[f(\mathbf{X}) g(\mathbf{X})] \approx \frac{1}{S} \sum_{s=1}^S f(\mathbf{X}(t_s)) g(\mathbf{X}(t_s)). \quad (52)$$

The final approximation highlights the motivation for this choice in a data-driven context: the inner product can be computed directly from samples  $\mathbf{X}(t_s) \sim \rho(\mathbf{x})$  drawn from the process. Using this inner product

$$[\Sigma]_{ij} \approx \frac{1}{S} \sum_{s=1}^S \phi_i(\mathbf{X}(t_s)) \phi_j(\mathbf{X}(t_s)), \quad [\Sigma_{\mathcal{K}_\tau}]_{ij} \approx \frac{1}{S} \sum_{s=1}^S \phi_i(\mathbf{X}(t_s)) \phi_j(\mathbf{X}(t_s + \tau)) \quad (53)$$

describe the correlation between measurement functions  $i$  and  $j$  measured instantaneously, or with delay  $\tau$ , respectively. Both  $\Sigma_{\mathcal{K}_\tau}$  and  $\Sigma^{-1}$  are simple to compute in biologically plausible online schemes.



## 6 Equivalence between CCA and coherent sets for Stable OU Process

This Section proves that the analytical approach to coherent set projection by finding the subdominant singular function is equivalent to the data-driven Galerkin projection [2], which is also equivalent to information bottleneck and canonical correlation analysis [7], for stable OU processes.

**Proposition 1.** *For stable OU processes, the theoretically predicted filters agree exactly with the CCA solution. That is,*

$$C_0^{-1}C_\tau C_0^{-1}C_{-\tau} = e^{A^\top \tau} \Sigma^{-1} e^{A\tau} \Sigma, \quad C_0^{-1}C_{-\tau} C_0^{-1}C_\tau = \Sigma^{-1} e^{A\tau} \Sigma e^{A^\top \tau}. \quad (54)$$

*Proof.* The solution of an OU process is given by

$$X_t = \int_{-\infty}^t e^{A(t-s)} \xi(s) ds. \quad (55)$$

We can then directly calculate the covariance matrices as follows.

$$\begin{aligned} C_0 &:= \mathbb{E}(X_t X_t^\top) \\ &= \int_{-\infty}^t \int_{-\infty}^t e^{A(t-s)} \mathbb{E}[\xi(s) \xi(s')^\top] e^{A^\top(t-s')} ds ds' \\ &= \int_{-\infty}^t \int_{-\infty}^t e^{A(t-s)} D \delta(s - s') e^{A^\top(t-s')} ds ds' \\ &= \int_{-\infty}^t e^{A(t-s)} D e^{A^\top(t-s)} ds \\ &= - \int_{-\infty}^t e^{A(t-s)} (A\Sigma + \Sigma A^\top) e^{A^\top(t-s)} ds \\ &= \int_{-\infty}^t \frac{d[e^{A(t-s)} \Sigma e^{A^\top(t-s)}]}{ds} ds \\ &= \Sigma, \end{aligned} \quad (56)$$

$$\begin{aligned} C_\tau &:= \mathbb{E}(X_t X_{t+\tau}^\top) \\ &= \int_{-\infty}^t \int_{-\infty}^{t+\tau} e^{A(t-s)} \mathbb{E}[\xi(s) \xi(s')^\top] e^{A^\top(t+\tau-s')} ds ds' \\ &= C_0 e^{A^\top \tau} \\ &= \Sigma e^{A^\top \tau}, \end{aligned} \quad (57)$$

$$\begin{aligned} C_{-\tau} &:= \mathbb{E}(X_t X_{t-\tau}^\top) \\ &= \int_{-\infty}^t \int_{-\infty}^{t-\tau} e^{A(t-s)} \mathbb{E}[\xi(s) \xi(s')^\top] e^{A^\top(t-\tau-s')} ds ds' \\ &= e^{A\tau} C_0 \\ &= e^{A\tau} \Sigma. \end{aligned} \quad (58)$$

From these equalities one can straightforwardly derive Eq. (54), showing that, for stable OU processes, our analytical method is equivalent to the data-driven method for finding the subdominant eigenfunction. Q.E.D.

## 7 Experimental Data Analysis

### 7.1 Dead Leaves Model Simulation

The plateaued stimulus used in Figure 2A was generated by concatenating  $\tanh$  functions with varying amplitude, time constant, and bias and sampled at unitary time steps. In Figure 2, the time on

the x axis is shown in the time steps of the dataset. To construct the lag vectors necessary to estimate the past and future temporal filters, the dataset was sampled by a sliding window 10 time steps long. The mean-centered lag vectors were processed by CCA.

## 7.2 Tufted/Mitral Cell Dataset

Data were taken from [8]. Briefly, anesthetized rats were presented with time-varying stimuli of multiple odors while tufted/mitral cells were recorded with extracellular tetrodes. For each odor-cell pair, an odor kernel best describing the response of a cell to a brief odor pulse was estimated. Kernels were estimated at 20 Hz with a total length of 2 seconds.

Kernels were parametrized as Gabor functions of the form

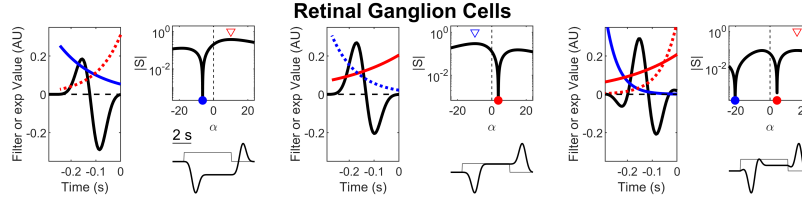
$$K(t) = ae^{(\frac{t-\tau}{w})^2} \cos(2\pi(f + t - \tau) + \phi)$$

with  $a$ ,  $\tau$ ,  $w$ ,  $f$ , and  $\phi$  as free parameters.

Here, we further set all kernel values with a magnitude below 0.005 to zero (to remove noise) and normalized all kernels to have unit  $\ell^2$  norm.

## 7.3 Retinal Ganglion Cell Dataset

Data were taken from [9]. Briefly, retinal ganglion cells were recorded using an electrode array during display of high contrast gaussian white noise. Spatio-temporal receptive fields (STRFs) were estimated using reverse correlation. For each cell, we extracted its temporal receptive field by computing a rank-1 decomposition using the SVD. This rank-1 approximation accounted for approximately 60% of the variance of the STRFs. Temporal kernels were estimated at 200 Hz with a total length of 0.5 seconds. Temporal receptive fields were then parameterized and normalized using the same procedure as above (Tufted/Mitral Cell Dataset).



**Figure S1:** Sample temporal receptive fields from three retinal ganglion cells in the salamander retina. As in Figure 3, the black lines indicate individual linear filters of different neurons, the solid colored lines indicate exponentials that are orthogonal to the filters, and the dotted colored lines indicate the exponentials that are most aligned with them. The bottom inset shows the convolution of the filter with a step pulse.

## 7.4 Identifying Orthogonal Exponentials

For each temporal receptive field in Figs. 2, 3, and S1, we numerically computed the cosine similarity between the filter and a normalized exponential.

The cosine similarity  $S$  between two column vectors was computed as

$$S(\mathbf{x}, \mathbf{y}) = \frac{\mathbf{x}^\top \mathbf{y}}{\|\mathbf{x}\| \|\mathbf{y}\|}$$

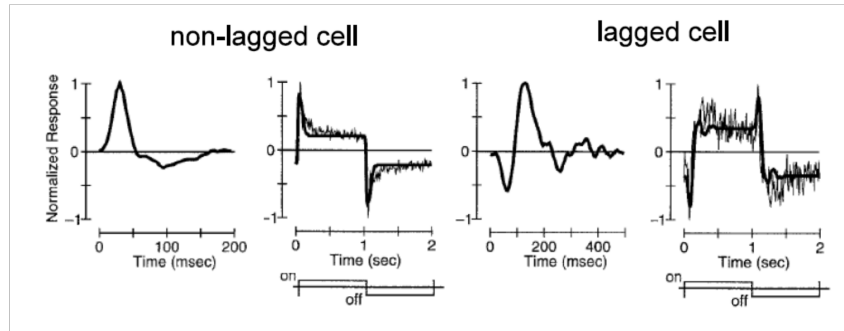
where  $\|\mathbf{u}\|$  is the  $\ell^2$  norm of the temporally discretized filter  $\mathbf{u} \in \mathbb{R}^n$ , where  $n = 40$  for the Tufted/Mitral cells and  $n = 100$  for the Retinal Ganglion cells.

Exponentials had the form

$$E(t) = ce^{\alpha t}$$

where  $c$  was chosen such that the  $L_2$  norm defined over the range of non-zero values of the temporal receptive field was 1. For tufted/mitral cells,  $\alpha$  was swept between -10 and 10. For retinal ganglion cells,  $\alpha$  was swept between -30 and 30. The zero-crossing of the cosine similarity as a function of  $\alpha$

was identified as the  $\alpha$  yielding an orthogonal exponential. The computation for all 1549 neurons (204 Mitral/Tufted, 1345 RGCs) took less than 1 minute on a standard PC laptop (Intel i7-1260P, 16GB RAM).



**Figure S2:** Sample temporal receptive fields and step responses from non-lagged and lagged cells of the cat LGN [10]. When comparing the temporal receptive fields with those in Fig. 3 and Fig. S1, note that the time axis is inverted. Step response lag accounts for the naming of the cells.

## References

- [1] Grigorios A Pavliotis. “Stochastic processes and applications”. In: *Texts in applied mathematics* 60 (2014).
- [2] Stefan Klus and Nataša Djurdjevic Conrad. “Dynamical systems and complex networks: A Koopman operator perspective”. In: *Journal of Physics: Complexity* 5.4 (2024), p. 041001.
- [3] Brian DO Anderson. “Reverse-time diffusion equation models”. In: *Stochastic Processes and their Applications* 12.3 (1982), pp. 313–326.
- [4] Predrag Cvitanović and Domenico Lippolis. “Knowing when to stop: How noise frees us from determinism”. In: *AIP Conference Proceedings*. Vol. 1468. 1. American Institute of Physics. 2012, pp. 82–126.
- [5] Jeffrey M Heninger, Domenico Lippolis, and Predrag Cvitanović. “Neighborhoods of periodic orbits and the stationary distribution of a noisy chaotic system”. In: *Physical Review E* 92.6 (2015), p. 062922.
- [6] Péter Koltai et al. “Optimal Data-Driven Estimation of Generalized Markov State Models for Non-Equilibrium Dynamics”. In: *Computation* 6.1 (Feb. 2018), p. 22. ISSN: 2079-3197. DOI: 10.3390/computation6010022.
- [7] Gal Chechik et al. “Information bottleneck for Gaussian variables”. In: *Advances in Neural Information Processing Systems* 16 (2003).
- [8] Priyanka Gupta, Dinu F Albeanu, and Upinder S Bhalla. “Olfactory bulb coding of odors, mixtures and sniffs is a linear sum of odor time profiles”. In: *Nature neuroscience* 18.2 (2015), pp. 272–281.
- [9] David B Kastner and Stephen A Baccus. “Coordinated dynamic encoding in the retina using opposing forms of plasticity”. In: *Nature Neuroscience* 14.10 (2011), pp. 1317–1322.
- [10] Daqing Cai, Gregory C Deangelis, and Ralph D Freeman. “Spatiotemporal receptive field organization in the lateral geniculate nucleus of cats and kittens”. In: *Journal of neurophysiology* 78.2 (1997), pp. 1045–1061.