Supplementary Information for

A Two-Dimensional Bayesian Continuous Attractor Neural Network for Robust Spatial State Estimation

Sohyun Kim[1] and Hyo-Sang Shin[1,2*]

[1] Cho Chun Shik Graduate School of Mobility, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, 34051, Republic of Korea.

[2] Faculty of Engineering and Applied Sciences, Cranfield University, Cranfield, MK43 0AL, UK.

*Corresponding author email: hyosangshin@kaist.ac.kr

Supplementary Note 1. Orthogonal projection for macroscopic neural dynamics

This Supplementary Note provides the detailed derivation of the macroscopic governing equations for the bump's center $\boldsymbol{\mu}_t$ and amplitude λ_t. We utilize the orthogonal projection method to extract these variables from the microscopic neural field. Starting from the combined neural field dynamics in equation (37) derived in the main text, the temporal evolution of the neural state $U(\mathbf{X}, t)$ is described as follows:

$$\begin{aligned}\frac{d\lambda_t}{dt}\frac{U(\mathbf{x},t)}{\lambda_t} - \frac{d\boldsymbol{\mu}_t}{dt}\cdot\nabla U(\mathbf{x},t) = {} & \left(\frac{-1 - G(\lambda_t) + I_{amp}/\lambda_t}{\tau}\right)U \\ & - \left(\frac{I_{amp}}{\tau\lambda_t}(\mathbf{z}_t - \boldsymbol{\mu}_t)\right)\cdot\nabla U \\ & + K_{rec}U^2 - 2w_{sym}(\mathbf{v}_t^{net}\cdot\nabla U)U.\end{aligned} \tag{S1}$$

Extracting the exact macroscopic variables from this nonlinear equation requires utilizing the symmetry properties of the Gaussian basis function. Under the manifold hypothesis, the neural state is parameterized as $U(\mathbf{x}, t) = \lambda_t \exp(-\|\mathbf{x} - \boldsymbol{\mu}_t\|^2/4\sigma_{bump}^2)$. Specifically, integrating the product of the even (symmetric) function U and the odd (anti-symmetric) function ∇U over the entire space $\mathbb{R}^2$ evaluates to exactly zero.

Pre-computed Gaussian integral identities. For the subsequent projection steps, the following Gaussian integral identities are pre-computed

$$\begin{gathered}\int U^2\,d\mathbf{x} = \lambda_t^2(2\pi\sigma_{bump}^2),\ \int U^3\,d\mathbf{x} = \lambda_t^3\left(\frac{4}{3}\pi\sigma_{bump}^2\right), \\ \int \nabla U \nabla U^T\,d\mathbf{x} = \lambda_t^2\left(\frac{\pi}{2}\right)\mathbf{I}_2,\ \int U\nabla U\nabla U^T\,d\mathbf{x} = \lambda_t^3\left(\frac{2\pi}{9}\right)\mathbf{I}_2,\end{gathered} \tag{S2}$$

where $\mathbf{I}_2$ denotes the 2×2 identity matrix.

Derivation of the bump's center dynamics. Utilizing these integral identities, the evolution equations for the macroscopic variables are derived by respectively projecting the neural field dynamics onto the basis functions ∇U and U. To isolate the translation dynamics, we multiply both sides of equation (S1) by ∇U and integrate over the spatial domain. Due to the symmetry, the integrals containing U and U^2 evaluate to zero, effectively eliminating the amplitude-related terms:

$$\begin{aligned}\left(\int \nabla U \nabla U^T \, d\mathbf{x}\right) \frac{d\boldsymbol{\mu}_t}{dt} = &- \frac{I_{amp}}{\tau \lambda_t} \left(\int \nabla U \nabla U^T \, d\mathbf{x}\right) (\mathbf{z}_t - \boldsymbol{\mu}_t) \\ &- 2w_{sym} \left(\int U \nabla U \nabla U^T \, d\mathbf{x}\right) \mathbf{v}_t^{net}.\end{aligned} \tag{S3}$$

Since the integral $\int \nabla U \nabla U^T \, d\mathbf{x}$ results in a scalar matrix, it can be factored out from both sides. Substituting the pre-computed integral values yields the temporal evolution of the bump's center:

$$\begin{aligned}\frac{d\boldsymbol{\mu}_t}{dt} &= 2w_{sym} \frac{\left(\int U \nabla U \nabla U^T d\mathbf{x}\right)}{\int \nabla U \nabla U^T d\mathbf{x}} \mathbf{v}_t^{net} + \frac{I_{amp}}{\tau \lambda_t} (\mathbf{z}_t - \boldsymbol{\mu}_t) \\ &= 2w_{sym} \frac{\lambda_t^3 (2\pi/9)}{\lambda_t^2 (\pi/2)} \mathbf{v}_t^{net} + \frac{I_{amp}}{\tau \lambda_t} (\mathbf{z}_t - \boldsymbol{\mu}_t) \\ &= \frac{8}{9} w_{sym} \lambda_t \mathbf{v}_t^{net} + \frac{I_{amp}}{\tau \lambda_t} (\mathbf{z}_t - \boldsymbol{\mu}_t).\end{aligned} \tag{S4}$$

Derivation of the bump's amplitude dynamics. Conversely, to isolate the amplitude dynamics, we multiply equation (S1) by U and integrate over the spatial domain. In this case, the gradient-related terms (∇U and $U \nabla U$) vanish due to their odd symmetry:

$$\frac{1}{\lambda_t} \frac{d\lambda_t}{dt} \int U^2 \, d\mathbf{x} = \left(\frac{-1 - G(\lambda_t) + I_{amp}/\lambda_t}{\tau}\right) \int U^2 \, d\mathbf{x} + K_{rec} \int U^3 \, d\mathbf{x}. \tag{S5}$$

Dividing both sides by $\int U^2 \, d\mathbf{x}$ and rearranging the terms gives the temporal derivative of the amplitude:

$$\begin{aligned}\frac{d\lambda_t}{dt} &= \frac{-1 - G(\lambda_t) + I_{amp}/\lambda_t}{\tau} \lambda_t + K_{rec} \frac{\int U^3 d\mathbf{x}}{\int U^2 d\mathbf{x}} \lambda_t \\ &= -\frac{1 + G(\lambda_t)}{\tau} \lambda_t + \frac{I_{amp}}{\tau} + K_{rec} \frac{\lambda_t^3 (\frac{4}{3} \pi \sigma_{bump}^2)}{\lambda_t^2 (2\pi \sigma_{bump}^2)} \lambda_t \\ &= -\frac{1 + G(\lambda_t)}{\tau} \lambda_t + \frac{I_{amp}}{\tau} + \frac{2}{3} K_{rec} \lambda_t^2.\end{aligned} \tag{S6}$$