A Moving Bump in a Continuous Manifold: A Comprehensive Study of the Tracking Dynamics of Continuous Attractor Neural Networks

3 authors:

C. C. Alan Fung
The Hong Kong University of Science and Technology
13 PUBLICATIONS   108 CITATIONS

SEE PROFILE

K. Y. Michael Wong
The Hong Kong University of Science and Technology
185 PUBLICATIONS   1,224 CITATIONS

SEE PROFILE

Si Wu
Beijing Normal University
118 PUBLICATIONS   1,866 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Integration and segregation in multisensory information processing View project

Spatial perception View project

A Moving Bump in a Continuous Manifold:

A Comprehensive Study of the Tracking Dynamics of Continuous Attractor Neural Networks

C. C. Alan Fung[1], K. Y. Michael Wong[1] and Si Wu[2,3]

[1]Department of Physics, Hong Kong University of Science and Technology,
Clear Water Bay, Hong Kong, China

[2]Department of Informatics, University of Sussex, Brighton, United Kingdom

[3]Lab of Neural Information Processing, Institute of Neuroscience, Shanghai, China

April 6, 2009

Abstract

Understanding how the dynamics of a neural network is shaped by the network structure, and consequently how the network structure facilitates the functions implemented by the neural system, is at the core of using mathematical models to elucidate brain functions. This study investigates the tracking dynamics of continuous attractor neural networks (CANNs). Due to the translational invariance of neuronal recurrent interactions, CANNs can hold a continuous family of stationary states. They form a continuous manifold in which the neural system is neutrally stable. We systematically explore how this property facilitates the tracking performance of a CANN, which is believed to have clear correspondence with brain functions. By using the wave functions of the quantum harmonic oscillator as the basis, we demonstrate how the dynamics of a CANN is decomposed into different motion modes, corresponding to distortions in the amplitude, position, width or skewness of the network state. We then develop a perturbative approach that utilizes the dominating movement of the network's stationary states in the state space. This method allows us to approximate the network dynamics up to an arbitrary accuracy depending on the order of perturbation used. We quantify the distortions of a Gaussian bump during tracking, and study their effects on the tracking performance. Results are obtained on the maximum speed for a moving stimulus to be trackable and the reaction time for the network to catch up with an abrupt change in the stimulus.

1

1 Introduction

The brain performs computation by updating its internal states according to dynamical rules that are determined by both the nature of external inputs and the structure of neural networks. A neural system acquires its network structure through either learning from experience, or inheriting from evolution or both. Its impact on network dynamics is twofold: on one hand, it determines the stationary states of the network that lead to associative memory; and on the other hand, it carves the landscape of the state space of the network as a whole, which may contribute to other cognitive functions, such as movement control, spatial navigation, population decoding and object categorization. Understanding how the dynamics of a neural network is shaped by its network structure is at the core of using mathematical models to elucidate brain functions (Dayan and Abbott 2001).

Associative memories, such as those based on the Hopfield model (Hopfield 1984; Domany, van Hemmen, and So have been intensively studied over the past decades. However, an equally important issue has not been investigated to the same extent, namely, how a large-scale structure of the state space of a neural system may facilitate brain functions. Recently, a type of recurrent network, called continuous attractor neural networks (CANNs), has received wide attention (Wilson and Cowan 1972; Amari 1977; Georgopoulos, Kalaska, Caminiti, and Massey 1982; Maunsell and Van Essen 1983; Funahashi, Bruce, and Goldman-Rakic 1989; Wilson and McNaughton 1993; Rolls, Robertson, and Georges-Franc Ben-Yishai, Lev Bar-Or, and Sompolinsky 1995; Zhang 1996; Seung 1996; Samsonovich and McNaughton 1997; Camperi and Wang 1998; Ermentrout 1998; Hansel and Sompolinsky 1998; Taube 1998; Deneve, Latham, and Po Wang 2001; Laing and Chow 2001; Wu, Amari, and Nakahara 2002; Stringer, Trappenberg, Rolls, and Aranjo 200 Brody, Romo, and Kepecs 2003; Erlhagen 2003; Gutkin, Pinto, and Ermentrout 2003; Renart, Song, and Wang 20 Trappenberg 2003; Folias and Bressloff 2004; Wu and Amari 2005; Chow and Coombes 2006; Miller 2006; Machens and Brody 2008; Wu, Hamaguchi, and Amari 2008). These models are particularly useful when modeling the encoding of continuous stimuli in neural systems. Different from other attractor models, CANNs have a distinctive feature that their neuronal interactions are translationally invariant. As a result, they can hold a family of stationary states that can be translated into each other without the need to overcome any barriers. In the continuum limit, these stationary states form a continuous manifold[1] in which the system is neutrally stable, and

---

[1]A manifold is a mathematical concept. Simply stated, it is a set of points with a coordinate system. Here, each stationary state of the network is a point in the manifold. As will be shown later, these stationary states have the shape of a Gaussian bump, and their peak positions define the coordinates.

the network state can translate easily when the external stimulus changes continuously. Beyond pure memory retrieval, this large-scale structure of the state space endows the neural system with a tracking capability. An illustration of the landscape of the state space of a CANN and how it leads to neutral stability of the network stationary states is shown in Fig. 1.

The tracking dynamics of a CANN has been theoretically studied by several authors in the literature (see, e.g.,(Ben-Yishai, Lev Bar-Or, and Sompolinsky 1995; Zhang 1996; Samsonovich and McNaughton 19 Xie, Hahnloser, and Seung 2002; Folias and Bressloff 2004; Wu and Amari 2005)). These studies have demonstrated clearly that a CANN has the capacity of tracking a moving stimulus continuously and that this tracking property can describe many brain functions well. Detailed rigorous analyses of the tracking behaviors of a CANN are still lacking, however, including, for instance, 1) the conditions under which a CANN can track a moving stimulus successfully; 2) how the network state is distorted during the tracking; and 3) how these distortions affect the tracking performance of a CANN. In this study, we systematically investigate these issues.

We use a simple, theoretically solvable model of a CANN as the working example. The stationary states of the network have a Gaussian shape, and the network dynamics can be described using the basis functions of quantum harmonic oscillators. These analytical results allow us to explore the tracking performances of a CANN in much more detail than in previous studies based on simulation observations. In particular, we demonstrate clearly how the dynamics of a CANN is decomposed into different modes corresponding to distortions in the height, position, width or skewness of the network bump states. Their contributions to the tracking performance are determined by the amplitudes of the corresponding eigenvalues of the neuronal interaction kernel. For a CANN, neutral stability implies that the mode of positional shift dominates the network dynamics and that other modes have high-order contributions. From this property, we develop a time-dependent perturbative approach to simplify the network dynamics. Geometrically, this corresponds to projecting the network dynamics on its dominating motion modes. The solution of the perturbation method is expressed in a simple closed form, and we can approximate the network dynamics up to an arbitrary accuracy depending on the order of perturbation used. With this method, we explore two tracking performances of a CANN, namely, the condition under which the network can successfully track a moving stimulus and the reaction time for the network to catch up with sudden changes in the stimulus. Both properties are associated with the unique dynamics of a CANN and can be tested in practice.

The idea of utilizing the special structure of a CANN to simplify its dynamics has also been

reported in other studies. Wu *et al* projected the dynamics of a CANN on its single dominating motion mode, the tangent of the attractor space, and simplify the network dynamics into a one-dimensional equation (Wu, Hamaguchi, and Amari 2008). Several other authors also proposed different approaches to simplifying the network dynamics (Ben-Yishai, Lev Bar-Or, and Sompolinsky 1995; Wennekers 2002; Renart, Song, and Wang 2003; Miller 2006). Compared with these studies, our contributions are that: 1) we elucidate the tracking behavior of a CANN in much more detail, particularly the effect of distortions in the network state on the tracking performance and the development of a perturbative method to systematically improve the description of the network dynamics; 2) we predict on the tracking behavior of a CANN, namely, the maximum speed for a moving stimulus to be trackable by the network (Hansel and Sompolinsky 1998) and the logarithmic nature of the reaction time; and 3) we also analytically investigate the tracking dynamics of a two-dimensional CANN. We expect that the mathematical framework developed in this work will have wide applications in the theoretical study of CANNs.

The organization of the paper is as follows. In Section 2, the unique dynamics of a CANN is explored, and we demonstrate how a CANN is decomposed into different modes. In Section 3, a simple method of projecting the network dynamics onto its single primary motion mode, i.e., the tangent of the manifold formed by all stationary states, is used to investigate the tracking performance of a CANN. In Section 4, a perturbative approach is developed to systematically improve the description of the tracking dynamics of the network. The effects of the shape distortions of the network state on the tracking dynamics are described. In Section 5, the dynamics of a two-dimensional CANN is investigated. In Section 6, we present the discussion and conclusions. The mathematical details are presented in the appendices. Preliminary results have been published in (Fung, Wong, and Wu 2008).

2   Intrinsic Dynamics of a CANN

Let us consider a one-dimensional continuous stimulus x encoded by an ensemble of neurons. For example, the stimulus may represent the direction of movement, the orientation or a general continuous feature of objects extracted by the neural system. We consider the range of possible values of the stimulus being much larger than the range of neuronal interactions. We can thus effectively take $x \in (-\infty, \infty)$.

Let $U(x, t)$ be the synaptic input at time t to the neurons whose preferred stimulus is x,
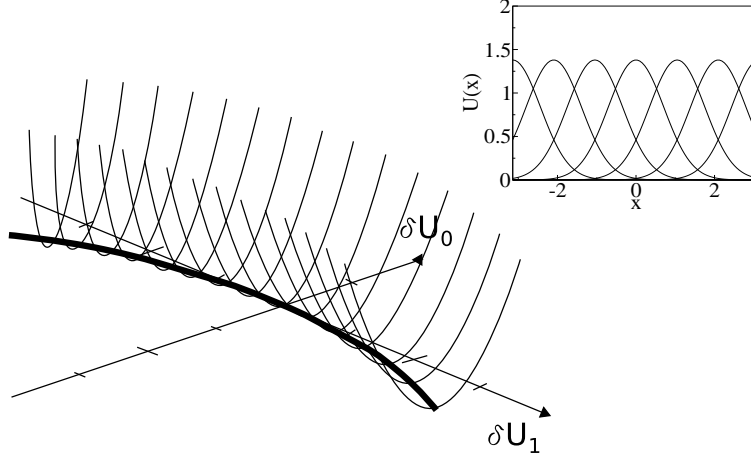
Figure 1: The energy landscape of the state space of a CANN. A canyon is formed by projecting the stationary states onto the subspace formed by δU_1 and δU_0, i.e., the position and the height distortions of the network states. Motion along the canyon corresponds to the positional shift of the bump (inset).

and let $r(x,t)$ be the firing rate of these neurons. $r(x,t)$ increases with the synaptic input, but saturates in the presence of global activity-dependent inhibition. A solvable model that captures these features is given by the divisive normalization (Deneve, Latham, and Pouget 1999; Wu, Amari, and Nakahara 2002):

$$r(x,t) = \frac{U(x,t)^2}{1 + k\rho \int_{-\infty}^{-\infty} dx' U(x',t)^2},\tag{1}$$

where ρ is the neural density and k is a small positive constant controlling the strength of global inhibition.

The dynamics of the synaptic input $U(x,t)$ is determined by the external input $I_{\text{ext}}(x,t)$, the network input from other neurons, and its own relaxation. It is given by

$$\tau\frac{\partial U(x,t)}{\partial t} = I_{\text{ext}}(x,t) + \rho \int_{-\infty}^{\infty} dx' J(x,x') r(x',t) - U(x,t),\tag{2}$$

where τ is a time constant, which is typically on the order of 1 ms (Ermentrout 1998; Gutkin, Pinto, and Ermentrou Wu, Hamaguchi, and Amari 2008), and $J(x,x')$ is the neural interaction from x' to x.

The key character of CANNs is the translational invariance of their neural interactions. In our solvable model, we choose Gaussian interactions with a range a, namely,

$$J(x,x') = \frac{A}{\sqrt{2\pi}a} \exp\left[-\frac{(x-x')^2}{2a^2}\right],\tag{3}$$

5

where A is a constant.

CANNs with other neural interactions and inhibition mechanisms have been studied in the literature (see, e.g., (Amari 1977; Ben-Yishai, Lev Bar-Or, and Sompolinsky 1995; Ermentrout 1998; Wang 2001)). Here, we choose this model for the advantage that it permits an analytical solution of the network dynamics. Nevertheless, the final conclusions of our model are qualitatively applicable to general cases.

We first consider the intrinsic dynamics of the CANN model in the absence of external stimuli. It is straightforward to check that, for $0 < k < k_c \equiv A^2\rho/(8\sqrt{2\pi}a)$, the network holds a continuous family of stationary states, which are (see Fig. 2A)

$$\tilde{U}(x|z) = U_0 \exp\left[-\frac{(x-z)^2}{4a^2}\right], \tag{4}$$

$$\tilde{r}(x|z) = r_0 \exp\left[-\frac{(x-z)^2}{2a^2}\right], \tag{5}$$

where $U_0 = [1 + (1 - k/k_c)^{1/2}]A/(4\sqrt{\pi}ak)$ and $r_0 = [1 + (1 - k/k_c)^{1/2}]/(2\sqrt{2\pi}ak\rho)$. These stationary states are translationally invariant among themselves and have a Gaussian shape with a free parameter z indicating their positions. z is therefore the coordinate of the manifold formed by all stationary states.

The stability of the Gaussian bumps can be studied by considering the dynamics of fluctuations. Consider the network state $U(x,t) = \tilde{U}(x|z) + \delta U(x,t)$. As derived in Appendix B, we have

$$\tau\frac{\partial}{\partial t}\delta U(x,t) = \int_{-\infty}^{\infty} dx' F(x,x'|z)\delta U(x',t) - \delta U(x,t), \tag{6}$$

where the interaction kernel $F(x,x'|z)$ is given by

$$F(x,x'|z) = \frac{2\rho U(x')}{B}\left[J(x,x') - k\rho\int_{-\infty}^{\infty} dx'' J(x,x'')r(x'')\right]. \tag{7}$$

We are interested in the eigenfunctions and eigenvalues of the kernel $F(x,x'|z)$. Since the kernel is not symmetric with respect to the permutation of x and x', we have to distinguish between its left and right eigenfunctions. To compute them, we need to choose a convenient set of basis functions. They should satisfy the boundary condition that they vanish when x approaches $\pm\infty$. A convenient choice is one that contains a basis function taking the shape of a Gaussian bump. Since the quantum ground state of the harmonic oscillator potential has a Gaussian shape, a natural choice of basis functions is the wave functions of quantum harmonic

6

oscillators (Griffiths 2004), namely,

$$v_n(x|z) = \frac{1}{\sqrt{(2\pi)^{1/2} a n! 2^n}} \exp\left[-\frac{(x-z)^2}{4a^2}\right] H_n\left(\frac{x-z}{\sqrt{2}a}\right), \tag{8}$$

where H_n are the nth-order Hermite polynomials for $n = 0, 1, 2, \ldots$, which can be expressed as the derivatives of the Gaussian function (Griffiths 2004)

$$v_n(x|z) = \frac{(-1)^n (\sqrt{2}a)^{n-1/2}}{\sqrt{\pi^{1/2} n! 2^n}} \exp\left[\frac{(x-z)^2}{4a^2}\right] \left(\frac{d}{dx}\right)^n \exp\left[-\frac{(x-z)^2}{2a^2}\right]. \tag{9}$$

This choice of basis functions has the further advantage that an abundance of mathematical properties are available for our analysis, as summarized in Appendix A.

Fig. 2B pictures the first four wave functions. We see that:

- The ground state of the quantum harmonic oscillator,

$$v_0(x|z) = \frac{\tilde{U}(x|z)}{U_0 \sqrt{(2\pi)^{1/2} a}}, \tag{10}$$

  corresponds to the stationary state of the network.

- The first excited state of the quantum harmonic oscillator,

$$v_1(x|z) = \frac{2a}{U_0 \sqrt{(2\pi)^{1/2} a}} \frac{\partial \tilde{U}(x|z)}{\partial z}, \tag{11}$$

  corresponds to the positional shift of the stationary state.

- The second and third excited states, $v_2(x|z)$ and $v_3(x|z)$, respectively correspond to the distortions in the width and skewness of the bump.

To express $\delta U(x,t)$ in terms of the eigenfunctions $v_n(x|z)$ using Eq. (6), we consider $F(x, x'|z)$ as an operator acting on $v_n(x|z)$. Since $v_n(x|z)$ is a basis, we have

$$\int_{-\infty}^{\infty} dx' F(x, x'|z) v_n(x'|z) = \sum_m v_m(x|z) F_{mn}, \tag{12}$$

where F_{mn} is the representation of the kernel $F(x, x'|z)$ in the basis $v_n(x|z)$, given by

$$F_{mn} = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dx' v_m(x|z) F(x, x'|z) v_n(x'|z). \tag{13}$$
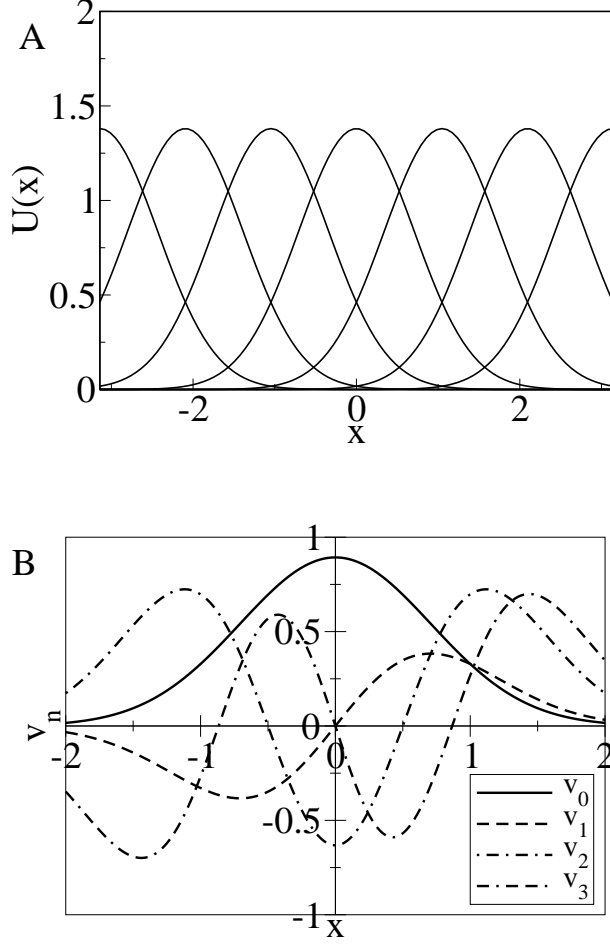
Figure 2: (A) The Gaussian-shaped stationary states of the network. (B) The first four basis functions of the quantum harmonic oscillators, representing distortions in the height, position, width and skewness of the Gaussian bump.

As derived in Appendix B for the particular choice of the output function (1) and the neural interactions (3), the elements of the matrix F_{mn} are given by

$$
F_{mn} = \begin{cases} 1 - \sqrt{1 - k/k_c}, & m = n = 0; \\[2mm] 2^{1-n} \sqrt{\dfrac{n!}{m!}} \dfrac{(-1)^{\frac{n-m}{2}}}{2^{\frac{n-m}{2}} \left(\dfrac{n-m}{2}\right)!}, & n - m \text{ being an even integer}; \\[2mm] 0, & \text{otherwise.} \end{cases} \tag{14}
$$

By using the matrix F_{mn}, we can calculate the eigenvalues and the first four right eigenfunctions of the kernel $F(x, x'|z)$, which are (see Appendix B)

$$
\lambda_0 = 1 - \sqrt{1 - k/k_c}, \tag{15}
$$

$$\lambda_n = 2^{1-n}, \quad \text{for } n \geq 1, \ldots, \tag{16}$$

$$u_0^R(x|z) = v_0(x|z), \tag{17}$$

$$u_1^R(x|z) = v_1(x|z), \tag{18}$$

$$u_2^R(x|z) = \frac{\sqrt{1/2}}{D_0}v_0(x|z) + \frac{1 - 2\sqrt{1 - k/k_c}}{D_0}v_2(x|z), \tag{19}$$

$$u_3^R(x|z) = \sqrt{\frac{1}{7}}v_1(x,z) + \sqrt{\frac{6}{7}}v_3(x,z), \tag{20}$$

where

$$D_0 = \left[\left(1 - 2\sqrt{1 - k/k_c}\right)^2 + 1/2\right]^{1/2}. \tag{21}$$

We note that $u_1^R(x|z) = v_1(x|z)$ and $\lambda_1 = 1$. Fig. 3A illustrates the first four right eigenfunctions of F.

The right eigenfunctions of F correspond to the various distortion modes of the bump. For example, $u_0^R(x|z)$ corresponds to the amplitude distortion of the bump, and $u_1^R(x|z)$ to the positional shift. Starting from $n = 3$, $u_n^R(x|z)$ are linear combinations of $v_k(x|z)$, where $k = n, n-2, \ldots$. Some of them may be dominated by a specific type of distortion, such as u_3^R in Eq. (20) being dominated by the skewness. However, generally speaking, various distortion features have been coupled in the eigenfunctions of the interaction kernel. To avoid confusion, we continue to use $v_n(x|z)$ as our basis functions in the rest of the paper.

To identify the contributions of $u_n^R(x|z)$ to the network dynamics, we express $\delta U(x,t)$ in (6) as a linear combination of $u_n^R(x|z)$, namely, $\delta U(x,t) = \sum_n \delta U_n(z,t)u_n^R(x|z)$. Using the orthonormality of the left and right eigenfunctions of $F(x, x'|z)$, we have

$$\delta U_n(z,t) = \int_{-\infty}^{\infty} dx \delta U(x,t)u_n^L(x|z). \tag{22}$$

Assume that the motion of the bump is slow, so that dz/dt becomes negligible in Eq. (6); as we shall see, this assumption is valid as long as the external stimulus is sufficiently weak. Then, the projection of Eq. (6) on the eigenfunctions become

$$\tau \frac{d}{dt}\delta U_n(z,t) = (\lambda_n - 1)\delta U_n(z,t). \tag{23}$$

Hence,

$$\delta U_n(z,t) = \delta U_n(z,0)\exp\left[-\frac{(1 - \lambda_n)t}{\tau}\right], \tag{24}$$

where $\delta U_n(z,0)$ is the initial value of the projection.

Eq. (24) tells us that, since $\lambda_1 = 1$, the bump distortion δU_1, which corresponds to the positional shift of the bump, is sustained under the network dynamics. This is the mathematical

expression of the neutral stability of the stationary states of the network. The eigenvalues for all other eigenfunctions are all less than one. Since these other modes of bump distortions decay exponentially with the time constant $\tau/(1 - \lambda_n)$, their contributions on the network dynamics are small when compared with that due to the positional shift. An intuitive illustration of the decomposition of the network dynamics is given in Fig. 3B.
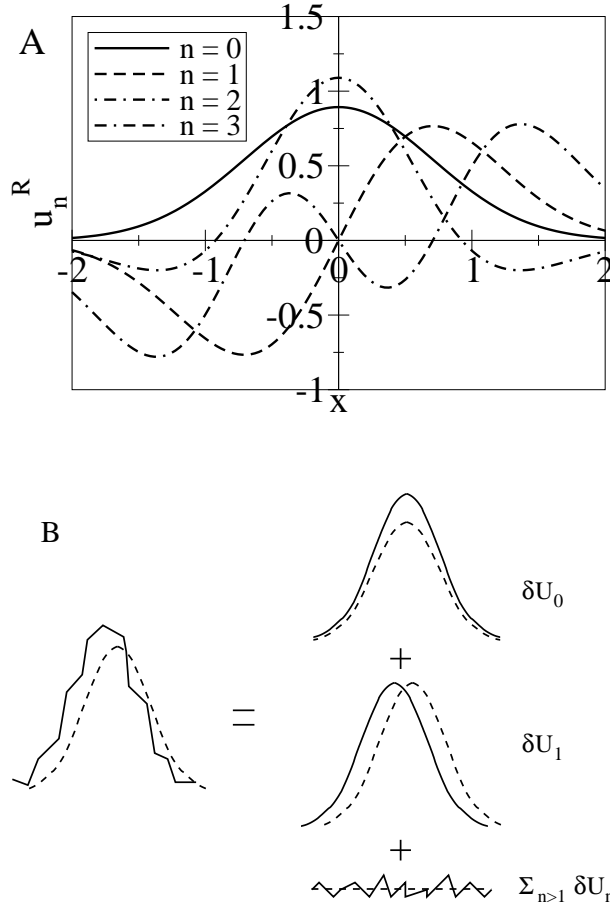

Figure 3: (A) The first four right eigenfunctions of the interaction kernel $F(x, x'|z)$ with $k/k_c = 1/2$. (B) The distortions of the network state are decomposed into three parts, namely, distortions in the height, position and all other shapes.

The fact that the dynamics of a CANN is dominated by only a few motion modes suggests that we may use this property to simplify the description of the network dynamics; that is, we can project the network dynamics onto its dominating modes. By this, we can analytically quantify the tracking bahviors of a CANN.

The Landscape of the State Space

It is instructive to illustrate the energy landscape in the state space of a CANN. In the dynamical system that we are studying, we do not have a Lyapunov function. This is evident from the fact that the interaction matrix F_{mn} is not symmetric. Nevertheless, at each peak position z, one can define an effective energy function $E(z) = \sum_n (1 - \lambda_n)[\delta U_n(z)]^2/2$, with $\delta U_n(z)$ being the projection of $U(x) - \tilde{U}(x|z)$ on the eigenfunction $u_n(x|z)$. Then, the dynamics in Eq. (6) or (23) can be locally described by the gradient descent of $E(z)$ in the space of $\delta U_n(z)$. $E(z)$ has the minimum value of zero when $\delta U_n(z) = 0$ for $n \neq 1$ and $\delta U_1(z)$ can take any value since $\lambda_1 = 1$. This corresponds to the picture that the bump's position shifts, but the shape of the bump remains unchanged. When z varies, the solution of $E(z) = 0$ traces a curve in the state space, facilitating the local gradient descent dynamics of the network, with the tangent of the curve at each point defining the direction of neutral stability. We can envisage this by a canyon surrounding the curve as shown in Fig. 1, and a small force along the tangent of the canyon can move the network state easily.

The External Input

To proceed, we need to define the external input. Without loss of generality, we choose

$$I_{\text{ext}}(x,t) = \alpha U_0 \exp\left[-\frac{(x - z_0)^2}{4a^2}\right] + \sigma\eta(x,t), \tag{25}$$

where α and σ indicate the strength of the signal and the noise, respectively. $\eta(x,t)$ is Gaussian white noise, satisfying $\langle\eta(x,t)\rangle = 0$ and $\langle\eta(x,t)\eta(x',t')\rangle = \delta(x - x')\delta(t - t')$. The signal term has the effect of driving the bump to the stimulus location z_0, and the noise term tends to shift the bump position randomly. Our results are qualitatively independent of the exact form of the external input. We may choose other forms for the external input. The main conclusions will be the same once the chosen external input has a unimodal shape and has the effect of driving the bump to the stimulus position.

The Setting of the Simulation Experiments

We will use simulation experiments to confirm the theoretical analysis. Here, we describe the setting of simulation experiments to be used below. The network consists of N neurons. Their preferred stimuli x are evenly distributed in the range $(-\pi, \pi]$ and satisfy the periodic condition, namely, $U(x,t) = U(x + 2\pi, t)$. In the simulation, the integration in the continuum limit is

changed to the summation accordingly, i.e., $\rho \int dx \to \sum_{i=1}^{N}$. We choose $a = 0.5 \ll 2\pi$, so that comparison with the theoretical analysis for $x \in (-\infty, \infty)$ holds.

3 Tracking Dynamics in the Weak Input Limit

We consider first the case when the external input is sufficiently small, i.e., $\alpha \ll 1$ and $\sigma \ll 1$. In this case, we can reasonably assume that the network dynamics is dominated by the positional shift of the bump, and we can thus ignore other motion modes, that is, $U(x,t) \approx \tilde{U}(x|z(t))$, with only the bump position $z(t)$ varying with time. As will be shown in Section 4, this actually corresponds to $n = 0$ perturbation in the general perturbative method.

After projecting the network dynamics, Eq. (2), on the basis functions $v_1(x|z(t))$, we obtain (see Appendix C)

$$\frac{dz}{dt} = -\frac{\alpha}{\tau}(z - z_0) \exp\left[-\frac{(z - z_0)^2}{8a^2}\right] + \sqrt{2D}\eta_1(t), \tag{26}$$

where $D = \sqrt{2}a\sigma^2/(U_0^2\tau^2\sqrt{\pi})$ is the diffusion constant, and $\eta_1(t)$ is the overlap of $\eta(x,t)$ with $v_1(x|z(t))$ given by

$$\eta_1(t) = \int_{-\infty}^{\infty} dx \eta(x,t) v_1(x|z(t)), \tag{27}$$

which satisfies

$$\langle \eta_1(t) \rangle = 0 \quad \text{and} \quad \langle \eta_1(t)\eta_1(t') \rangle = \delta(t - t'). \tag{28}$$

This equation represents a one-dimensional Ornstein-Uhlenbeck process for the motion of the bump position in the state space (Tuckwell 1988). Its meaning is straightforward: the first term on the right-hand side of the equation represents the contribution of the stimulus, whose effect is to pull the bump to the stimulus position z_0; the second term represents noise, which tends to shift the bump position randomly. This equation agrees with Eq. (3.7) in (Wu, Hamaguchi, and Amari 2008), which considers $|z - z_0| \ll a$ and hence $e^{-(z-z_0)^2/8a^2} \approx 1$.

The distribution of the bump position can be calculated. The pull term in Eq. (26) can be written as the gradient of a potential,

$$V(z) = -\frac{4a^2\alpha}{\tau} \exp\left[-\frac{(z - z_0)^2}{8a^2}\right], \tag{29}$$

so that Eq. (26) can be written as a Langevin equation,

$$\frac{dz}{dt} = -\frac{\partial V}{\partial z} + \sqrt{2D}\eta(t). \tag{30}$$

12

This implies that z satisfies a Boltzmann distribution (Ma 1976),

$$P(z) \sim \exp\left[-\frac{V(z)}{D}\right]. \tag{31}$$

Since the pull is maximum at $z = z_0 \pm 2a$, two noise regimes can be identified. In the low noise regime, where $\langle (z - z_0)^2 \rangle \ll 4a^2$, the bump position is effectively trapped in a parabolic potential, whereas in the high noise regime with $\langle (z - z_0)^2 \rangle \gg 4a^2$, the pull vanishes and the motion of the bump resembles a random walk with $\langle [z(t) - z_0]^2 \rangle = 2Dt$ loosely coupled to the stimulus.

However, simulation results show that Eq. (31) underestimates the variance of the bump position. This is because we have considered only the overlap of $\eta(x, t)$ with $v_1(x|z(t))$. As will be shown in the perturbative approach, overlaps with other components cannot be neglected.

By using the simplified dynamics in Eq. (26), we explore two tracking behavior of the network, namely, the tracking of a moving stimulus and the catching up of an abrupt change in the stimulus.

3.1   The Condition for Successful Tracking

Suppose that the stimulus is moving at a constant velocity v and the noise is low and negligible. The dynamical equation becomes identical to Eq. (26) after the transients, with z_0 replaced by vt. Denoting the lag of the bump behind the stimulus as $s = z_0 - z$, we have

$$\begin{aligned}
\frac{ds}{dt} &= \frac{dz_0}{dt} - \frac{dz}{dt} \\
&= v - \frac{\alpha s}{\tau} \exp\left[-\frac{s^2}{8a^2}\right] \\
&= v - g(s).
\end{aligned} \tag{32}$$

The size of the lag s is determined by two competing factors: the first term represents the movement of the stimulus, which tends to enlarge the lag; and the second term represents the collective effects of the neuronal recurrent interactions, which tends to shorten the lag. The tracking is achieved when these two forces match each other, i.e., $v = g(s)$; otherwise, s diverges.

Fig. 4A shows that the function $g(s)$ is concave and has the maximum value of $g_{\max} = 2\alpha a/(\tau \sqrt{e})$ at $s = 2a$. This means that if $v > g_{\max}$, the network is unable to track the stimulus. Thus, $g_{\max}$ defines the maximum trackable speed of a moving stimulus to the network. Notably, $g_{\max}$ increases with the strength of the external signal α and the range of neuronal recurrent interactions a. This is reasonable since it is the neuronal interactions that induce the movement

of the bump. $g_{\max}$ decreases with the time constant of the network τ, as this reflects the responsiveness of the network to external inputs.

On the other hand, for $v < g_{\max}$, there is a stable and unstable fixed point, denoted as s_1 and s_2 respectively, in Fig. 4A. When the initial lag is less than s_2, it will converge to s_1. Otherwise, the tracking of the stimulus will be lost. Intuitively, this means that the propagation of neuronal activities in the network needs to be quick enough to compensate for the initial lag behind the stimulus. Figs. 4B and 4C show that the analytical results of Eq. (32) agree well with the simulation results.

3.2 The Reaction Time

Another tracking behavior we study is the response to a stimulus experiencing an abrupt change. We measure how long it takes for the network to catch up with this change, i.e., the reaction time. Suppose that the network has reached a steady state at $t < 0$, and the stimulus position jumps from 0 to z_0 suddenly at $t = 0$. This is the typical setting in experiments studying mental rotation behavior (Georgopoulos, Taira, and Lukashin 1993; Ben-Yishai, Lev Bar-Or, and Sompolinsky 1995; Wu and Amari 2005). The reaction time can be obtained by integrating Eq. (26) with the initial value $z(t = 0) = 0$.

When the noise is low and the jump size $|z_0|$ is much smaller than the range a of neuronal interactions, we can derive an analytical expression for the reaction time. In this case, Eq. (26) is approximated as

$$\tau \frac{dz}{dt} \approx \alpha(z_0 - z). \tag{33}$$

We define the reaction time T as the time it takes for the bump to move to a small threshold distance θ from the stimulus. T is calculated to be

$$T \approx \frac{\tau}{\alpha} \ln\left(\frac{|z_0|}{\theta}\right). \tag{34}$$

This reveals that with small jump sizes, the reaction time of a CANN increases logarithmically with the jump size.

Fig. 5 compares the simulation results with our theoretical predictions based on Eq. (26). We see that with small jump sizes, they agree very well with each other, and the reaction time indeed has a logarithmic nature in its increase with the jump size. However, we also notice that with large jump sizes, there is considerable discrepancy between the simulation results and

the theoretical predictions. This is because we have only considered the positional shift in the network dynamics. To improve the results, more motion modes must be included.

4   Tracking Dynamics: A Comprehensive Study

In this section, we develop a time-dependent perturbation approach to systematically improve the description of the network dynamics. Since the network dynamics is primarily dominated by the positional shift of the bump, with high-order distortions from other motion modes, we consider the network states with the following form,

$$U(x,t) = \tilde{U}(x|z(t)) + \sum_{n=0}^{\infty} a_n(t)v_n(x|z(t)), \tag{35}$$

where the bump position $z(t)$ and the coefficients $\{a_n(t)\}$ specify the network state.

The external input can be similarly expressed as

$$I_{\text{ext}}(x,t) = \sum_{n=0}^{\infty} I_n(t)v_n(x|z), \tag{36}$$

with $I_n = \int dx I_{\text{ext}}(x,t)v_n(x|z)$.

By substituting Eqs. (35) and (36) into Eq. (2), and making use of the orthonormality and completeness of $v_n(x|z)$, we obtain the perturbative equation for $a_n(t)$ as (see Appendix D)

$$\left(\frac{d}{dt} + \frac{1-\lambda_n}{\tau}\right)a_n = \frac{I_n}{\tau} - \left[U_0\sqrt{(2\pi)^{1/2}a}\delta_{n1} + \sqrt{n}a_{n-1} - \sqrt{n+1}a_{n+1}\right]\frac{1}{2a}\frac{dz}{dt}$$
$$+ \frac{1}{\tau}\sum_{r=1}^{\infty}\sqrt{\frac{(n+2r)!}{n!}}\frac{(-1)^r}{2^{n+3r-1}r!}a_{n+2r}. \tag{37}$$

The center-of-mass position $z(t)$ is determined from the self-consistent condition,

$$z(t) = \frac{\int_{-\infty}^{\infty} dxU(x,t)x}{\int_{-\infty}^{\infty} dxU(x,t)}. \tag{38}$$

By combining Eqs.(37) and (38) and denoting the product of the positive integers $n, n-2, \ldots$ as $n!!$, we obtain an equation for the bump velocity (see Appendix D),

$$\frac{dz}{dt} = \frac{2a}{\tau}\frac{I_1 + \sum_{n=3,\text{odd}}^{\infty}\sqrt{\frac{n!!}{(n-1)!!}}I_n + a_1}{U_0\sqrt{(2\pi)^{1/2}a} + \sum_{n=0,\text{even}}^{\infty}\sqrt{\frac{(n-1)!!}{n!!}}a_n}. \tag{39}$$

Eqs. (37) and (39) are the master equations of the perturbative method. For the perturbation order specified by n, we include the coefficients from a_0 up to a_n, and the input components

15

from I_0 up to I_n, and at least I_1, which is the lowest order of input that drives the dynamics. By choosing different orders of perturbation, we can approximate the network dynamics to different accuracies.

The components of the external input in $v_n(x|z)$ are calculated to be (see Appendix E)

$$I_n(t) = \alpha U_0 \exp\left[-\frac{(z_0 - z)^2}{8a^2}\right] \left(\frac{z_0 - z}{2a}\right)^n \sqrt{\frac{(2\pi)^{1/2}a}{n!}} + \sigma\eta_n(t), \tag{40}$$

where the noise term $\eta_n(t)$ satisfies $\langle\eta_n(t)\rangle = 0$ and $\langle\eta_m(t)\eta_n(t')\rangle = \delta_{mn}\delta(t - t')$.

4.1 The Weak Input Limit

We first compare the perturbative approach with the weak input limit by ignoring all distortion modes of the bump; that is, we set $a_n = 0$ for all n. Due to the readiness of the bump to shift its position, the lowest order of the external stimulus to be included is I_1 according to Eq. (39). Thus, in the $n = 0$ perturbation, Eq. (39) becomes

$$\frac{dz}{dt} = \frac{2a}{\tau} \frac{I_1}{U_0\sqrt{(2\pi)^{1/2}a}}. \tag{41}$$

After substituting I_1 from Eq. (40) into Eq. (41), we obtain Eq. (26), which is equivalent to the network dynamics in the limit of weak inputs.

4.2 The $n = 0$ Perturbation

We further consider the $n = 0$ perturbation. This amounts to taking into account the effects of the change in the bump height on the network dynamics, apart from the positional shift. Perturbation up to $n = 0$ yields

$$\left(\frac{d}{dt} + \frac{1 - \lambda_0}{\tau}\right)a_0 = \frac{\alpha U_0}{\tau}\sqrt{(2\pi)^{1/2}a}\exp\left[-\frac{(z_0 - z)^2}{8a^2}\right] + \frac{\eta_0}{\tau}, \tag{42}$$

$$\frac{dz}{dt} = \frac{\alpha U_0\sqrt{(2\pi)^{1/2}a}(z_0 - z)\exp\left[-\frac{(z_0-z)^2}{8a^2}\right] + 2a\sigma\eta_1}{\tau(U_0\sqrt{(2\pi)^{1/2}a} + a_0)}. \tag{43}$$

With the initial condition $a_0 = 0$ at $t = 0$, the solution of a_0 from Eq. (42) reduces to

$$a_0(t) = \int_0^t \frac{dt'}{\tau}\exp\left[-\frac{1 - \lambda_0}{\tau}(t - t')\right]\left\{\alpha U_0\sqrt{(2\pi)^{1/2}a}\exp\left[-\frac{(z_0 - z)^2}{8a^2}\right] + \eta_0(t')\right\}. \tag{44}$$

Substituting Eq. (44) into Eq. (43), we have

$$\frac{dz}{dt} = \left\{\frac{\alpha}{\tau}(z_0 - z)\exp\left[-\frac{(z_0 - z)^2}{8a^2}\right] + \sqrt{2D}\eta_1\right\}R(t)^{-1}, \tag{45}$$

16

where

$$R(t) = 1 + \alpha \int_0^t \frac{dt'}{\tau} \exp\left[-\frac{1-\lambda_0}{\tau}(t-t') - \frac{(z_0 - z(t'))^2}{8a^2}\right] + \frac{\xi_0(t)}{U_0\sqrt{(2\pi)^{1/2}a}}, \tag{46}$$

and $\xi_0(t)$ is a stochastic variable given by

$$\xi_0(t) = \sigma \int_0^t \frac{dt'}{\tau} \exp\left[-\frac{(1-\lambda_0)(t-t')}{\tau}\right] \eta_0(t'). \tag{47}$$

$R(t)$ can be interpreted to be the ratio of the bump height in the presence of the external stimulus to its weak input limit. Compared with Eq. (26), Eq. (45) has an extra term $R(t)^{-1}$. $R(t)^{-1}$ approaches 1 in the limit of weak inputs, but, in general, $R(t)^{-1} < 1$, implying that the increase in the amplitude of the bump has the effect of slowing down the tracking dynamics. In the following two applications, we consider the low noise limit with $\sigma = 0$.

4.2.1 Tracking a Moving Stimulus

Similiar to the derivation of Eq. (32), we can determine the dynamical equation for the bump lag s after the transient period. From Eq. (45),

$$\frac{ds}{dt} = v - \frac{\alpha s}{\tau} \exp\left(-\frac{s^2}{8a^2}\right) R(t)^{-1}, \tag{48}$$

where $R(t)$ is given by

$$R(t) = 1 + \alpha \int_{-\infty}^t \frac{dt'}{\tau} \exp\left[-\frac{1-\lambda_0}{\tau}(t-t') - \frac{s(t')^2}{8a^2}\right]. \tag{49}$$

If the stimulus is not too weak, s reaches its steady state value in a short duration. The integral can then be computed precisely, yielding

$$R(t) = 1 + \frac{\alpha}{1-\lambda_0} \exp\left[-\frac{s^2}{8a^2}\right]. \tag{50}$$

Compared with Eq. (32), Eq. (48) has a correction term, $R(t)^{-1} < 1$, representing the effect of the bump height distortion.

4.2.2 Tracking an Abrupt Change in the Stimulus

The dynamical equation in this case is identical to Eq. (45) with the initial value $z(t = 0) = 0$. To compute the correction term $R(t)^{-1}$, however, we need to consider that the bump height is initially at a steady value determined by the steady-state response of the network to the stimulus

at $z = 0$. This means that Eq. (44) needs to be replaced by

$$
\begin{aligned}
a_0(t) &= \frac{\alpha U_0 \sqrt{(2\pi)^{1/2}a}}{1 - \lambda_0} \exp\left[-\frac{(1 - \lambda_0)t}{\tau}\right] \\
&+ \alpha U_0 \sqrt{(2\pi)^{1/2}a} \int_0^t \frac{dt'}{\tau} \exp\left[-\frac{1 - \lambda_0}{\tau}(t - t') - \frac{(z_0 - z)^2}{8a^2}\right],
\end{aligned} \tag{51}
$$

and the factor $R(t)$ in Eq. (45) is replaced by

$$
R(t) = 1 + \frac{\alpha}{1 - \lambda_0} \exp\left[-\frac{(1 - \lambda_0)t}{\tau}\right] + \alpha \int_0^t \frac{dt'}{\tau} \exp\left[-\frac{1 - \lambda_0}{\tau}(t - t') - \frac{(z_0 - z(t'))^2}{8a^2}\right]. \tag{52}
$$

Indeed, $R(t)$ represents the change in height during the movement of bump. The second and third terms respectively contribute to the heights near the initial and final positions (see Fig. 5A).

Fig. 6 compares the simulation results and the theoretical predictions based on the perturbation approximation of Eqs. (45) and (52). We see that compared with the weak input limit, the result is improved a little bit.

4.3 The $n = 5$ Perturbation

By increasing the order of perturbation, we get an increasingly better approximation of the network dynamics. We observe that the improvement is btter when the order of perturbation n increases by 1 to an odd integer. This is because odd distortion modes are antisymmetric, which better approximate the bump's shape distortion when the position of the external stimulus is biased towards one side of the bump.

Below, we present the results of the $n = 5$ perturbation, which correspond to taking into account distortions in the position, height, width, skewness and the fourth- and fifth-order distortions of the bump.

By including perturbations up to $n = 5$, Eqs. (37) and (39) become

$$
\left(\frac{d}{dt} + \frac{1 - \lambda_0}{\tau}\right) a_0 = \frac{I_0}{\tau} + \frac{a_1}{2a}\frac{dz}{dt} - \frac{\sqrt{2}}{4\tau}a_2 + \frac{\sqrt{6}}{32\tau}a_4, \tag{53}
$$

$$
\frac{da_1}{dt} = \frac{I_1}{\tau} - \left(\sqrt{(2\pi)^{1/2}a}U_0 + a_0 - \sqrt{2}a_2\right)\frac{1}{2a}\frac{dz}{dt} - \frac{\sqrt{6}}{8\tau}a_3 + \frac{\sqrt{30}}{64\tau}a_5, \tag{54}
$$

$$
\left(\frac{d}{dt} + \frac{1}{2\tau}\right) a_2 = \frac{I_2}{\tau} - \frac{\sqrt{2}a_1 - \sqrt{3}a_3}{2a}\frac{dz}{dt} - \frac{\sqrt{3}}{8\tau}a_4, \tag{55}
$$

$$
\left(\frac{d}{dt} + \frac{3}{4\tau}\right) a_3 = \frac{I_3}{\tau} - \frac{\sqrt{3}a_2 - \sqrt{4}a_4}{2a}\frac{dz}{dt} - \frac{\sqrt{5}}{16\tau}a_5, \tag{56}
$$

$$
\left(\frac{d}{dt} + \frac{7}{8\tau}\right) a_4 = \frac{I_4}{\tau} - \frac{\sqrt{4}a_3 - \sqrt{5}a_5}{2a}\frac{dz}{dt}, \tag{57}
$$

18

$$\frac{dz}{dt} = \frac{2a\left(I_1 + \sqrt{\frac{3}{2}}I_3 + \sqrt{\frac{15}{8}}I_5 + a_1\right)}{\tau\left(\sqrt{(2\pi)^{1/2}a}U_0 + a_0 + \sqrt{\frac{1}{2}}a_2 + \sqrt{\frac{3}{8}}a_4\right)},$$ (58)

and a_5 is determined by the condition $a_1 + \sqrt{3/2}a_3 + \sqrt{15/8}a_5 = 0$. Fig. 7 compares the theoretical predictions with the simulation results. They agree very well even when the jump sizes are large.

4.4 Shape Distortion

Besides the reaction time, it is instructive to study the distortion in the bump's shape during tracking. As shown in Fig. 8, there is an abrupt change in the height, position, width and skewness immediately after the external stimulus is abruptly removed at $t = 0$. $a_0(t)$ approaches 0 roughly, as if there were no external stimulus, and there is even a slight overshoot. $a_1(t)$ also drops abruptly, showing that the peak position lags behind the center of mass, an indication that the bump is abruptly distorted. At the same time, distortions in the width and skewness increase abruptly, showing that the bump is suddenly pulled by the external stimulus from its new position. After the initial abrupt changes, the distorted components undergo rather smooth changes on the journey to the new position. Finally, when the bump approaches its destination, the height returns to its equilibrium value, while the other components approach 0.

We also compare the perturbation and simulation results in Fig. 8. The agreement is very good. Discrepancies arise from linearizing the interaction term in going from Eq. (2) to (6). Their effects are mainly seen in the moments immediately after the abrupt change in the external stimulus.

Figure 9 shows the distorted bump in the center-of-mass frame during tracking. During the process, the lowering of the height, the retarding of the peak position, the broadening of the width, and the skewing due to the attraction of the newly positioned stimulus are clearly visible.

5 Dynamics of a Two-Dimensional CANN

We can extend our analysis to a two-dimensional(2D) CANN. Consider a neural ensemble encoding a 2D continuous stimulus $\mathbf{x} = (x_1, x_2)$, with its components $x_1 \in (-\infty, \infty)$ and $x_2 \in (-\infty, \infty)$. We denote by $U(\mathbf{x})$ the synaptic input to neurons having the preferred stimulus

$\mathbf{x}$ and $r(\mathbf{x})$ the corresponding firing rate. For a solvable model, we consider the network whose dynamics is governed by

$$\tau \frac{\partial U(\mathbf{x}, t)}{\partial t} = I_{\text{ext}}(\mathbf{x}, t) + \rho \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} d\mathbf{x}' J(\mathbf{x}, \mathbf{x}') r(\mathbf{x}', t) - U(\mathbf{x}', t), \tag{59}$$

$$r(\mathbf{x}, t) = \frac{U(\mathbf{x}, t)^2}{1 + k\rho \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} d\mathbf{x}' U(\mathbf{x}', t)^2}, \tag{60}$$

where τ is the time constant and ρ the neural density, and the recurrent interactions are chosen to be

$$J(\mathbf{x}, \mathbf{x}') = \frac{A}{2\pi a^2} \exp\left[-\frac{(\mathbf{x} - \mathbf{x}')^2}{2a^2}\right], \tag{61}$$

where $(\mathbf{x} - \mathbf{x}')^2 = (x_1 - x_1')^2 + (x_2 - x_2')^2$ is the Euclidean distance between $\mathbf{x}$ and $\mathbf{x}'$.

It can be checked that when $I_{\text{ext}}(\mathbf{x}, t) = 0$ and $0 < k < k_c \equiv A^2 \rho/(32\pi a^2)$, the network holds a continuous family of stationary states given by

$$\tilde{U}(\mathbf{x}|\mathbf{z}) = U_0 \exp\left[-\frac{(\mathbf{x} - \mathbf{z})^2}{4a^2}\right], \tag{62}$$

$$\tilde{r}(\mathbf{x}|\mathbf{z}) = r_0 \exp\left[-\frac{(\mathbf{x} - \mathbf{z})^2}{2a^2}\right], \tag{63}$$

where $U_0 = [1 + (1 - k/k_c)^{1/2}]A/(8\pi a^2 k)$ and $r_0 = [1 + (1 - k/k_c)^{1/2}]/(4\pi a^2 k\rho)$. The parameter $\mathbf{z} = (z_1, z_2)$ is a free variable, indicating the location of the bump. We note the differences in $J(\mathbf{x}, \mathbf{x}')$, k_c, U_0 and r_0 when compared with their values in the 1D case.

We analyze the stability of the bump states, and consider the network state $U(\mathbf{x}, t) = \tilde{U}(\mathbf{x}|\mathbf{z}) + \delta U(\mathbf{x}, t)$. After linearizing Eq. (59) at $\tilde{U}(\mathbf{x}|\mathbf{z})$, we get

$$\tau \frac{\partial}{\partial t} \delta U(\mathbf{x}, t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} d\mathbf{x}' F(\mathbf{x}, \mathbf{x}'|\mathbf{z}) \delta U(\mathbf{x}', t) - \delta U(\mathbf{x}, t), \tag{64}$$

where the interaction kernel is given by

$$\begin{aligned}
F(\mathbf{x}, \mathbf{x}'|\mathbf{z}) &= \frac{2}{a^2 \pi} \exp\left[-\frac{(\mathbf{x} - \mathbf{x}')^2}{2a^2}\right] \exp\left[-\frac{(\mathbf{x}' - \mathbf{z})^2}{4a^2}\right], \\
&\quad - \frac{1 + \sqrt{1 - k/k_c}}{2\pi a^2} \exp\left[-\frac{(\mathbf{x} - \mathbf{z})^2}{4a^2}\right] \exp\left[-\frac{(\mathbf{x}' - \mathbf{z})^2}{4a^2}\right]. \tag{65}
\end{aligned}$$

We can construct the right eigenfunction of $F(\mathbf{x}, \mathbf{x}'|\mathbf{z})$ by using the product of the eigenfunctions $u_n(x|z)$ in the 1D case, i.e.,

$$u_{m,n}(\mathbf{x}|\mathbf{z}) = u_m(x_1|z_1) u_n(x_2|z_2), \quad m, n = 0, 1, 2, \ldots. \tag{66}$$

The corresponding eigenvalues are given by (see Appendix F)

$$\lambda_{0,0} = \lambda_0 \tag{67}$$

$$\lambda_{m,0} = \lambda_m, \quad m \neq 0, \tag{68}$$

$$\lambda_{0,n} = \lambda_n, \quad n \neq 0, \tag{69}$$

$$\lambda_{m,n} = \lambda_m \lambda_n, \quad m \neq 0, n \neq 0, \tag{70}$$

where λ_m are the eigenvalues in the 1D case given by Eqs. (15) and (16).

The eigenfunction $u_{0,0}(\mathbf{x}|\mathbf{z})$ corresponds to the amplitude mode of the 2D bump state of the network. $u_{1,0}(\mathbf{x}|\mathbf{z})$ corresponds to the displacement of the bump position along the coordinate z_1 and $u_{0,1}(\mathbf{x}|\mathbf{z})$ the positional shift along z_2. A linear combination of $u_{1,0}(\mathbf{x}|\mathbf{z})$ and $u_{0,1}(\mathbf{x}|\mathbf{z})$, i.e., $c_1 u_{1,0}(\mathbf{x}|\mathbf{z}) + c_2 u_{0,1}(\mathbf{x}, \mathbf{z})$, with c_1 and c_2 being constants and satisfying $c_1^2 + c_2^2 = 1$, corresponds to the positional shift of the bump along the $\mathbf{c} = (c_1, c_2)$ direction. Interestingly, the eigenvalue for $c_1 u_{1,0}(\mathbf{x}|\mathbf{z}) + c_2 u_{0,1}(\mathbf{x}, \mathbf{z})$ is 1, indicating that the bump is neutrally stable in the two-dimensional space $\mathbf{z}$.

Analogous to the 1D case, we can develop a perturbation method to solve the network dynamics, e.g., by using the 2D wave functions, $v_{m,n}(\mathbf{x}|\mathbf{z}) = v_n(x_1|z_1) v_n(x_2|z_2)$, of the quantum harmonic oscillators as the basis functions. A more comprehensive treatment of the problem involves transforming the basis functions to polar coordinates; it will be presented elsewhere. Here, as an illustration, we will present only the result for the weak input limit in the rectilinear coordinates, which corrsponds to projecting the network dynamics on a single displacement mode. We choose the external input to be

$$I_{\text{ext}}(\mathbf{x}, t) = \alpha U_0 \exp\left[-\frac{(\mathbf{x} - \mathbf{z}^0)^2}{4a^2} \right] + \sigma \eta(\mathbf{x}, t), \tag{71}$$

where $\mathbf{z}^0$ represents the stimulus location.

Suppose the bump position is $\mathbf{z}(t)$ at time t. Then, the stimulus will pull the bump towards the location $\mathbf{z}^0$. In the limit of weak inputs, the network dynamics is dominated by the positional shift of the bump along the straight line through $\mathbf{z}^0$. As derived in Appendix F,

$$\frac{d(\mathbf{z} - \mathbf{z}^0)}{dt} = \frac{\alpha}{\tau}(\mathbf{z} - \mathbf{z}^0) \exp\left[-\frac{(\mathbf{z} - \mathbf{z}^0)^2}{8a^2} \right] + \sqrt{2D}\eta(t), \tag{72}$$

where $D = \sigma^2/(U_0^2 \tau^2 \pi)$. This equation describes how the distance between the bump and the stimulus varies with time. It has the same form as Eq. (26) except that $z - z^0$ is replaced by $\mathbf{z} - \mathbf{z}^0$. Thus, the network exhibits the same dynamical properties as observed in the 1D case.

Fig. 10 compares the theoretical predictions and the simulation results of the reaction time of the 2D CANN. As expected, they agree with each other very well with small jump sizes.

6   Discussion and Conclusions

The idea that the landscape of the energy function of a CANN has a canyon structure, which facilitates the tracking capacity of a neural system has long been recognized (see, e.g., (Amari 1977; Hansel and Sompolinsky 1998; Brody, Romo, and Kepecs 2003; Wu and Amari 2005)). However, a detailed and rigorous analysis of this property has been lacking. The main challenge here is to solve the network dynamics analytically, elucidating how the shape of the network state is affected during the tracking. In this study, by using a simple analytically solvable model, we systematically investigate this issue. There are two parts to our main results. First, we show that the dynamics of a CANN can be decomposed into different motion modes, corresponding to distortions in the height, position, width, skewness and other higher-order distortions in the shape of the bump states. We demonstrate that with a CANN, it is the positional shift that dominates the network dynamics, with other motion modes having high-order contributions. Second, we develop a time-dependent perturbation approach to approximate the network dynamics. Geometrically, this corresponds to projecting the network dynamics onto its dominating motion modes. Simulation results confirm that our method works very well. Two interesting phenomena in the tracking performance of a CANN are observed, namely, the maximum trackable speed and the logarithmic reaction time of the network. Since both properties are associated with the unique dynamics of a CANN, they may be tested in practice and serve as clues for checking the application of a CANN in neural systems.

To facilitate an analytical solution and hence to describe the network dynamics clearly, we have used a simple firing-rate-based model with the inhibition effect in the form of divisive normalization, and the generality of our results may thus be a concern. In particular, we should consider whether the mechanism of divisive normalization is plausible, whether tracking behavior can be observed in models with different inhibitory mechanisms, and how the perturbation method developed in this paper can be used in other types of networks.

First, we argue that a neural system can have resources to implement divisive normalization (Grossberg 1988; Heeger 1993; Deneve, Latham, and Pouget 1999). Let us consider a network in which all excitatory neurons are connected to a pool of inhibitory neurons. If the

time constant of inhibitory neurons is much smaller than that of excitatory neurons, and if these inhibitory neurons inhibit the activity of the excitatory neurons in a uniform shunting way, then the effect of divisive normalization can be achieved.

Second, even if there exists uncertainty of the model, the main conclusions of tracking behavior in this work are applicable to general cases. This is because our calculation is based on the fact that the dynamics of a CANN is largely dominated by the positional shift of the network states, a property coming from the translational invariance of the neuronal interactions, rather than from the inhibition mechanism. To see this point clearly, let us consider the dynamics of a network having the same form as Eq. (2), but the exact form of the inihibition mechansim remains unspecified. We consider $J(x - x')$ and $r(x)$ that are properly chosen (e.g., $J(x - x')$ is of the Mexcian-hat shape and $r(x)$ a sigmoid function of the inputs), so that the network holds a continuous family of stationary states, denoted as $\tilde{U}(x|z)$ and $\tilde{r}(x|z)$, with z representing the bump position. This implies that

$$\tilde{U}(x|z) = \rho \int_{-\infty}^{\infty} dx' J(x - x')\tilde{r}(x'|z). \tag{73}$$

By differentiating both sides of the above equation with respect to z, we get

$$\begin{aligned}\frac{\partial \tilde{U}(x|z)}{\partial z} &= \rho \int_{-\infty}^{\infty} dx' \int_{-\infty}^{\infty} dx'' J(x - x')\frac{\partial \tilde{r}(x'|z)}{\partial \tilde{U}(x''|z)}\frac{\partial \tilde{U}(x''|z)}{\partial z} \\ &= \int_{-\infty}^{\infty} dx'' F(x, x''|z)\frac{\partial \tilde{U}(x''|z)}{\partial z}.\end{aligned} \tag{74}$$

Thus, the mode of the positional shift of the bump, $\partial \tilde{U}(x|z)/\partial z$, is the eigenfunction of the interaction kernel F with the eigenvalue equal to 1. We should observe similar tracking behavior of the network, once $J(x-x')$ and $r(x)$ are properly chosen to ensure that all the other eigenvalues are less than one.

Third, the perturbation method developed in this paper can be used to analyze other types of networks with tracking behavior. Consider appropriate choices of $J(x - x')$ and $r(x)$ that can hold a continuous family of bump states centered at position z (for example, $J(x - x')$ having a Mexican-hat shape and $r(U)$ a sigmoid function). Then, the dynamics of the bump can be described by Eq. (6), with

$$F(x, x'|z) = \int_{-\infty}^{\infty} dx' J(x - x')\frac{\partial r(\tilde{U}(x'|z))}{\partial \tilde{U}(x'|z)}. \tag{75}$$

Using the basis functions of the quantum harmonic oscillator, we can obtain the matrix elements F_{mn} from Eq. (13). While elegant expressions such as those in Eq. (14) may not be available in

the general case, we can nevertheless compute F_{mn} numerically. Perturbation dynamics similar to Eqs. (37) and (39) can be worked out following Appendix D. Indeed, we have considered networks with Mexican-hat-shaped interactions. Tracking dynamics with a logarithmic reaction time and a maximum trackable speed is observed, and the convergence of the perturbation method is equally remarkable. These results will be reported elsewhere.

The tracking dynamics of a CANN has also been studied by other authors. In particular, Zhang proposed a mechanism of using asymmetrical recurrent interactions to drive the bump, so that the shape distortion would be minimized (Zhang 1996). Xie *et al* further proposed a double ring network model that uses the inputs from the vestibular system in mammals to achieve these asymmetrical interactions (Xie, Hahnloser, and Seung 2002). This mechansim works well for the head-direction system, but it is unclear if it can be achieved in other neural systems. For instance, in the visual and hippocampal systems, it is often assumed that the bump movement is directly driven by external inputs (see, e.g., (Samsonovich and McNaughton 1997; Berry II, Brivanlou, Jordon, and Meister 1999; Fu, Shen, and Dan 2001)), and the distortion of the bump is inevitable (indeed the bump distortions in (Berry II, Brivanlou, Jordon, and Meister 1999; Fu, Shen, and Dan 2001) are associated with visual perception). The contribution of this study is that we quantify how the distortion of the bump's shape affects the network tracking performance, and we obtain the maximum trackable speed of the network. We hope that this study will enrich our knowledge about the tracking dynamics of a CANN, which may differ in different brain functions. Apart from the tracking capacity, recent experimental observations have suggested that CANNs may have other important roles in neural cognitive functions. For example, the highly structured state space of a CANN may provide a neural basis of encoding the categorization relationships of objects (Jastorff, Kourtzi, and Giese 2006; Graf, Wichmann, Bulthoff, and Scholkopf 2006). It is quite possible that the distance between two memory states in the canyon defines the perceptual similarity between the two objects. We will explore these issues in the future, and we expect that the theoretical model and approaches developed in this work will provide valuable tools.

Acknowledgment

UK.

Appendix A: The Hermite Polynomials

Since the Hermite polynomials are frequently used in the calculations of this paper, here we summarize their main properties for the convenience of the reader.

The nth-order Hermite polynomial is defined as

$$H_n(x) = (-1)^n \exp(x^2) \left(\frac{d}{dx}\right)^n \exp(-x^2). \tag{76}$$

The first five Hermite polynomials are given by $H_0(x) = 1$, $H_1(x) = 2x$, $H_2(x) = 4x^2 - 2$, $H_3(x) = 8x^3 - 12x$, $H_4(x) = 16x^4 - 48x^2 + 12$. For the analysis in this paper, it is instructive to note that the Hermite polynomials satisfy the following properties.

- Recursion relation:

$$H_{n+1}(x) = 2xH_n(x) - H_n'(x), \tag{77}$$

$$H_n'(x) = 2nH_{n-1}(x), \tag{78}$$

$$H_n(x+y) = \sum_{k=0}^{n} \binom{n}{k} H_k(x)(2y)^{n-k}. \tag{79}$$

- Contour integral representation:

$$H_n(x) = n! \oint \frac{dt}{2\pi i t^{n+1}} e^{2tx - t^2}, \tag{80}$$

with the contour encircling the origin.

Appendix B: The Interaction Kernel $F(x, x'|z)$

Linearizing Eq. (2) at $\tilde{U}(x|z)$, we arrive at

$$\tau \frac{\partial}{\partial t} \delta U(x) = \rho \int_{-\infty}^{\infty} dx' J(x, x') \left[\frac{2U(x')}{B} \delta U(x') \right.$$
$$\left. - \frac{U(x')^2}{B^2} \int_{-\infty}^{\infty} dx'' 2k\rho U(x'')\delta U(x'') \right] - \delta U(x), \tag{81}$$

where we have dropped the dependence on t for brevity, and $B = 1 + k\rho \int_{-\infty}^{\infty} dx' \tilde{U}(x')^2 = \rho A U_0 / \sqrt{2}$. Interchanging the dummy variables x' and x'' in the last term, we obtain Eqs. (6)

and (7), with

$$
\begin{aligned}
F(x, x'|z) &= \frac{2}{a\sqrt{\pi}} \exp\left[-\frac{(x-x')^2}{2a^2}\right] \exp\left[-\frac{(x'-z)^2}{4a^2}\right] \\
&\quad - \frac{1+\sqrt{1-k/k_c}}{\sqrt{2\pi}a} \exp\left[-\frac{(x-z)^2}{4a^2}\right] \exp\left[-\frac{(x'-z)^2}{4a^2}\right].
\end{aligned}
\tag{82}
$$

By substituting the expressions of $U(x)$, $J(x,x')$ and $r(x)$, we get Eq. (7).

To calculate the matrix elements F_{mn} in Eq. (13), we denote

$$
g_n(x|z) = \int_{-\infty}^{\infty} dx' F(x, x'|z) v_n(x'|z).
\tag{83}
$$

From Eqs. (7) and (8),

$$
\begin{aligned}
g_n(x|z) &= \frac{(-1)^n 2(\sqrt{2}a)^{n-\frac{1}{2}}}{a\sqrt{\pi^{3/2}n!2^n}} \int_{-\infty}^{\infty} dx' \exp\left[-\frac{(x-x')^2}{2a^2}\right] \left(\frac{d}{dx'}\right)^n \exp\left[-\frac{(x'-z)^2}{2a^2}\right] \\
&\quad - \frac{(1+\sqrt{1-k/k_c})(-1)^n(\sqrt{2}a)^{n-\frac{1}{2}}}{a\sqrt{\pi^{3/2}n!2^{n+1}}} \exp\left[-\frac{(x-z)^2}{4a^2}\right] \\
&\quad \times \int_{-\infty}^{\infty} dx' \left(\frac{d}{dx'}\right)^n \exp\left[-\frac{(x'-z)^2}{2a^2}\right].
\end{aligned}
\tag{84}
$$

The second term vanishes for $n \geq 1$. Hence, by evaluating the two integrals explicitly for $n = 0$, we have

$$
g_0(x|z) = (1 - \sqrt{1-k/k_c}) v_0(x|z).
\tag{85}
$$

Making use of the orthonormality of the basis $v_n(x|z)$, we obtain the first line of Eq. (14).

Next, we consider the case $n \geq 1$, focusing on the first term.

Utilizing the identity $(-d/dz)^n e^{-(x-z)^2/2a^2} = (d/dx)^n e^{-(x-z)^2/2a^2}$, we have

$$
g_n(x|z) = \frac{(-1)^n 2(\sqrt{2}a)^{n-\frac{1}{2}}}{a\sqrt{\pi^{3/2}n!2^n}} \left(-\frac{d}{dz}\right)^n \int_{-\infty}^{\infty} dx' \exp\left[-\frac{(x-x')^2}{2a^2}\right] \exp\left[-\frac{(x'-z)^2}{2a^2}\right].
\tag{86}
$$

Evaluating the integral explicitly, we find that

$$
g_n(x|z) = \frac{(-1)^n 2(\sqrt{2}a)^{n-\frac{1}{2}}}{\sqrt{\pi^{1/2}n!2^n}} \left(\frac{d}{dx}\right)^n \exp\left[-\frac{(x-z)^2}{4a^2}\right].
\tag{87}
$$

We are now ready to derive F_{mn} from $g_n(x|z)$:

$$
F_{mn} = \int_{-\infty}^{\infty} dx v_m(x|z) g_n(x|z).
\tag{88}
$$

From Eqs. (8) and (86),

$$
\begin{aligned}
F_{mn} &= 2 \frac{(-1)^m(\sqrt{2}a)^{m-\frac{1}{2}}}{\sqrt{\pi^{1/2}m!2^m}} \frac{(-1)^n(\sqrt{2}a)^{n-\frac{1}{2}}}{\sqrt{\pi^{1/2}n!2^n}} \\
&\quad \times \int_{-\infty}^{\infty} dx \exp\left[\frac{(x-z)^2}{4a^2}\right] \left(\frac{d}{dx}\right)^m \exp\left[-\frac{(x-z)^2}{2a^2}\right] \left(\frac{d}{dx}\right)^n \exp\left[-\frac{(x-z)^2}{4a^2}\right].
\end{aligned}
\tag{89}
$$

26

Using the definition of the Hermite polynomials in Eq. (76), this simplfies to

$$F_{mn} = 2\frac{(-1)^m(\sqrt{2}a)^{m+n-1}}{\sqrt{\pi}m!n!2^{m+n}}\frac{1}{(2a)^n}\int_{-\infty}^{\infty}dx H_n\left(\frac{x-z}{2a}\right)\left(\frac{d}{dx}\right)^m\exp\left[-\frac{(x-z)^2}{2a^2}\right].\qquad(90)$$

Applying integration by parts m times yields

$$F_{mn} = 2\frac{(\sqrt{2}a)^{m+n-1}}{\sqrt{\pi}m!n!2^{m+n}}\frac{1}{(2a)^n}\int_{-\infty}^{\infty}dx\exp\left[-\frac{(x-z)^2}{2a^2}\right]\left(\frac{d}{dx}\right)^m H_n\left(\frac{x-z}{2a}\right).\qquad(91)$$

Applying Eq. (78) m times yields, for $n \geq m$,

$$F_{mn} = 2\frac{(\sqrt{2}a)^{m+n-1}}{\sqrt{\pi}m!n!2^{m+n}}\frac{1}{(2a)^n}\frac{n!}{a^m(n-m)!}\int_{-\infty}^{\infty}dx\exp\left[-\frac{(x-z)^2}{2a^2}\right]H_{n-m}\left(\frac{x-z}{2a}\right),\qquad(92)$$

and $F_{mn} = 0$ otherwise.

Introducing the contour integral representation in Eq. (80) leads to

$$F_{mn} = \frac{\sqrt{2}\sqrt{n!}}{2^n a\sqrt{\pi m!}}\int_{-\infty}^{\infty}dx\exp\left[-\frac{(x-z)^2}{2a^2}\right]\oint\frac{dt}{2\pi it^{n-m+1}}\exp\left[-t^2+t\left(\frac{x-z}{a}\right)\right].\qquad(93)$$

Interchanging the order of integrating x and t, we have

$$F_{mn} = 2^{1-n}\sqrt{\frac{n!}{m!}}\oint\frac{dt}{2\pi it^{n-m+1}}\exp\left(-\frac{t^2}{2}\right).\qquad(94)$$

According to the residue theoreom in complex analysis, the contour integral can be reduced to the coefficient of t^{n-m} in the Laurent expansion of $\exp(-t^2/2)$, which is equal to $(-1/2)^{(n-m)/2}/[(n-m)/2]!$ when $n - m$ is even and 0 otherwise. Hence, we have derived F_{mn} in the second line of Eq. (14).

Other than the above cases, all other elements F_{mn} are zero. Hence we have derived Eq. (14) completely.

As an illustration, we write down the elements of $\mathbf{F}$ in the first four rows and columns,

$$\mathbf{F} = \begin{pmatrix} 1-\sqrt{1-k/k_c} & 0 & -\frac{\sqrt{2}}{4} & 0 & \ldots \\ 0 & 1 & 0 & -\frac{\sqrt{6}}{8} & \ldots \\ 0 & 0 & \frac{1}{2} & 0 & \ldots \\ 0 & 0 & 0 & \frac{1}{4} & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \end{pmatrix}.\qquad(95)$$

Since all elements of the matrix $\mathbf{F}$ in the lower half of the diagonal line vanish, the eigenvalues of $\mathbf{F}$ are given by the diagonal elements, namely, Eqs. (15) and (16). It is straightforward to check that the first four right eigenvectors of F_{mn} are given by

$$\mathbf{q}_0 = (1,0,0,\ldots)^T,\qquad(96)$$

$$\mathbf{q}_1 = (0, 1, 0, \ldots)^T, \tag{97}$$

$$\mathbf{q}_2 = (\frac{\sqrt{1/2}}{D_0}, 0, \frac{1 - 2\sqrt{1 - k/k_c}}{D_0}, 0, \ldots)^T, \tag{98}$$

$$\mathbf{q}_3 = (0, \sqrt{\frac{1}{7}}, 0, \sqrt{\frac{6}{7}}, 0, \ldots)^T, \tag{99}$$

where D_0 is given in Eq. (21) and all the omitted elements are zero. The right eigenfunctions of $F(x, x'|z)$, with eigenvalues λ_n, are given by

$$u_n^R(x|z) = \sum_l q_{nl} v_l(x|z). \tag{100}$$

This can be verified by considering the inner product

$$\int_{-\infty}^{\infty} dx' F(x, x'|z) u_k^R(x'|z) = \int_{-\infty}^{\infty} dx' \sum_{mn} v_m(x|z) F_{mn} v_n(x'|z) \sum_l q_{kl} v_l(x'|z). \tag{101}$$

Making use of the orthonormality of $v_l(x'|z)$, we have

$$\int_{-\infty}^{\infty} dx' F(x, x'|z) u_k^R(x'|z) = \sum_{ml} v_m(x|z) F_{ml} q_{kl}. \tag{102}$$

Since $\mathbf{q}_k$ is an eigenvector of $\mathbf{F}$ with eigenvalue λ_k,

$$\int_{-\infty}^{\infty} dx' F(x, x'|z) u_k^R(x'|z) = \lambda_k \sum_m v_m(x|z) q_{km} = \lambda_k u_k^R(x|z). \tag{103}$$

The first four right eigenfunctions of $\mathbf{F}$ are presented in Eqs. (17)-(20).

Appendix C: The Network Dynamics in the Weak Input Limit

We let $U(x, t) = \tilde{U}(x|z(t))$ and multiply both sides of Eq. (2) by $v_1(x|z(t))$, and then we integrate over x. Since

$$\frac{\partial \tilde{U}(x|z(t))}{\partial t} = \frac{U_0 \sqrt{(2\pi)^{1/2} a}}{2a} \frac{dz}{dt} v_1(x|z(t)), \tag{104}$$

the left-hand side of the resultant equation becomes

$$L.H.S. = \frac{\tau U_0 \sqrt{(2\pi)^{1/2} a}}{2a} \frac{dz}{dt}. \tag{105}$$

On the right hand side of Eq.(2), the first term arises from the external input, consisting of both signal and noise components according to Eq. (25). The signal component, when projected onto the component $v_1(x|z(t))$, leads to

$$R.H.S._{\text{signal}} = \frac{\alpha U_0}{\sqrt{(2\pi)^{1/2} a}} \int_{-\infty}^{\infty} dx \exp\left[-\frac{(x - z_0)^2}{4a^2}\right] \left(\frac{x - z}{a}\right) \exp\left[-\frac{(x - z)^2}{4a^2}\right]. \tag{106}$$

Explicit integration yields

$$R.H.S._{\text{signal}} = \alpha U_0 \sqrt{(2\pi)^{1/2} a} \frac{z_0 - z}{2a} \exp\left[-\frac{(z_0 - z)^2}{8a^2}\right]. \tag{107}$$

The noise component leads to

$$R.H.S._{\text{noise}} = \sigma \eta_1(t), \tag{108}$$

where η_1 is given in Eq. (27) and satisfies Eq. (28).

The second and third terms of Eq. (2) vanish owing to the fact that $\tilde{U}(x|z(t))$ is the stationary state solution of Eq. (2) in the weak input limit, independent of the bump position $z(t)$.

Combining the above results, we get Eq. (26).

Appendix D: The Perturbation Method

D.1 The Perturbation Equation

Using Eq. (8), we have

$$\frac{d}{dt} v_n(x|z) = \frac{1}{\sqrt{(2\pi)^{1/2} a n! 2^n}} \exp\left[-\frac{(x-z)^2}{4a^2}\right] \left[\frac{x-z}{2a^2} H_n\left(\frac{x-z}{\sqrt{2}a}\right) - \frac{1}{\sqrt{2}a} H_n'\left(\frac{x-z}{\sqrt{2}a}\right)\right] \frac{dz}{dt}. \tag{109}$$

By virtue of Eq. (77) and (78),

$$\frac{d}{dt} v_n(x|z) = \left[\sqrt{n+1} v_{n+1}(x|z) - \sqrt{n} v_{n-1}(x|z)\right] \frac{1}{2a} \frac{dz}{dt}. \tag{110}$$

Similarly, since $\tilde{U}(x|z) = U_0 \sqrt{(2\pi)^{1/2} a} v_0(x|z)$,

$$\frac{\partial}{\partial t} \tilde{U}(x|z) = U_0 \sqrt{(2\pi)^{1/2} a} v_1(x|z) \frac{1}{2a} \frac{dz}{dt}. \tag{111}$$

The neuronal interaction and relaxation terms cancel each other at the steady state. Hence, in determining the dynamics, only the linearized terms need to be considered,

$$\rho \int_{-\infty}^{\infty} dx' J(x, x') r(x'|z) - U(x|z) \approx \sum_n a_n \sum_m v_m(x|z) F_{mn} - \sum_n a_n v_n(x|z). \tag{112}$$

Substituting Eqs. (110-112) into Eq. (2), we obtain

$$\tau \sum_n \frac{da_n}{dt} v_n + \frac{\tau}{2a} \left\{ U_0 \sqrt{(2\pi)^{1/2} a} v_1 + \sum_n a_n \left[\sqrt{n+1} v_{n+1} - \sqrt{n} v_{n-1}\right] \right\} \frac{dz}{dt}$$
$$= \sum_n a_n \sum_m v_m F_{mn} - \sum_n a_n v_n + \sum_n I_n v_n. \tag{113}$$

Making use of the orthonormality and completeness of $v_n(x|z)$, the coefficients in Eq. (113) can be equated term by term. Hence, we get

$$\tau \frac{da_n}{dt} + \frac{\tau}{2a} \left[U_0 \sqrt{(2\pi)^{1/2} a} \delta_{n1} + \sqrt{n} a_{n-1} - \sqrt{n+1} a_{n+1} \right] \frac{dz}{dt} = \sum_k F_{nk} a_k - a_n + I_n. \quad (114)$$

By further using Eq. (14), the above equation is re-organized as Eq. (37).

D.2 The Velocity Equation

The peak position $z(t)$ of the bump is determined from the self-consistent condition in Eq. (38). Without loss of generality, we consider $z = 0$ and substitute the expression of $v_n(x|z)$ into Eq. (38),

$$\sum_{n=1,\text{odd}} \frac{a_n}{\sqrt{n! 2^n}} \int_{-\infty}^{\infty} dx \exp\left(-\frac{x^2}{4a^2}\right) x H_n\left(\frac{x}{\sqrt{2}a}\right) = 0. \quad (115)$$

Integrating by parts and using the identity in Eq. (78), we have

$$\sum_{n=1,\text{odd}} \frac{n a_n}{\sqrt{n! 2^n}} \int_{-\infty}^{\infty} dx \exp\left(-\frac{x^2}{4a^2}\right) H_{n-1}\left(\frac{x}{\sqrt{2}a}\right) = 0. \quad (116)$$

The integral can be performed by using the contour integral representation of the Hermite polynomial in Eq. (80),

$$\sum_{n=1,\text{odd}} \frac{n! a_n}{\sqrt{n! 2^n}} \int_{-\infty}^{\infty} dx \oint \frac{dt}{2\pi i t^n} \exp\left(-t^2 - \frac{x^2}{4a^2} + \frac{2tx}{\sqrt{2}a}\right) = 0. \quad (117)$$

Interchanging the order of integration yields

$$\sum_{n=1,\text{odd}} \frac{n! a_n}{\sqrt{n! 2^n}} \oint \frac{dt}{2\pi i t^n} \exp(t^2) = 0. \quad (118)$$

The contour integral is equal to the coefficient of t^{n-1} in the series expansion of $\exp(t^2)$, which is $[(n-1)/2]!$. Hence,

$$\sum_{n=1,\text{odd}} \sqrt{\frac{n!!}{(n-1)!!}} a_n = 0. \quad (119)$$

Multiplying both sides of Eq. (37) by $\sqrt{n!!/(n-1)!!}$, and summing over odd n, the time derivatives of a_n disappear, leaving behind

$$0 = \frac{1}{\tau} \sum_{n=1,\text{odd}} \sqrt{\frac{n!!}{(n-1)!!}} I_n$$
$$- \left[U_0 \sqrt{(2\pi)^{1/2} a} + \sum_{n=1,\text{odd}} \sqrt{\frac{n!!}{(n-1)!!}} \left(\sqrt{n} a_{n-1} - \sqrt{n+1} a_{n+1}\right) \right] \frac{1}{2a} \frac{dz}{dt}$$
$$+ \frac{1}{\tau} \sum_{n=1,odd} \sqrt{\frac{n!!}{(n-1)!!}} \sum_{r=0}^{\infty} \sqrt{\frac{(n+2r)!}{n!}} \frac{(-1)^r}{2^{n+3r-1} r!} a_{n+2r}. \quad (120)$$

30

Collecting terms in the summations, we have

$$\left\{ U_0 \sqrt{(2\pi)^{1/2}a} + \sum_{n=0,\text{even}} \left[\sqrt{\frac{(n+1)!!}{(n)!!}}\sqrt{n+1} - \sqrt{\frac{(n-1)!!}{(n-2)!!}}\sqrt{n} \right] a_n \right\} \frac{1}{2a}\frac{dz}{dt}$$

$$= \frac{1}{\tau}\sum_{n=1,\text{odd}} \sqrt{\frac{n!!}{(n-1)!!}}(I_n - a_n) + \frac{1}{\tau}\sum_{m=1,\text{odd}} \sqrt{m!} \left[\sum_{r=0}^{(m-1)/2} \frac{(-1)^r}{2^{m+r-1}(m-2r-1)!!r!} \right] a_m. \quad (121)$$

Note that for odd m, $(m-2r-1)!! = 2^{(m-2r-1)/2}[(m-1)/2-r]!$, so that the summation over r in the last term is proportional to the binomial expansion of $[1+(-1)]^{(m-1)/2}$, which is equal to 1 for $m = 1$, and vanishes otherwise. Therefore, Eq. (39) is obtained.

Appendix E: The Components of the External Input

We first calculate the component of the signal term in Eq. (25),

$$I_n = \alpha U_0 \int_{-\infty}^{\infty} dx \exp\left[-\frac{(x-z_0)^2}{4a^2} \right] v_n(x|z). \quad (122)$$

After substituting Eq. (8) and completing square in the exponential argument, we have

$$I_n = \frac{\alpha U_0}{\sqrt{(2\pi)^{1/2}an!2^n}} \exp\left[-\frac{(z_0-z)^2}{8a^2} \right] \int_{-\infty}^{\infty} dx \exp\left[-\frac{(x-(z_0+z)/2)^2}{2a^2} \right] H_n\left(\frac{x-z}{\sqrt{2}a} \right). \quad (123)$$

Using the identity in Eq. (79), denoting $(z_0+z)/2$ as $\bar{z}$, and rewriting the integrand in terms of the basis functions, we obtain

$$\begin{aligned} I_n &= \frac{\alpha U_0}{\sqrt{(2\pi)^{1/2}an!2^n}} \exp\left[-\frac{(z_0-z)^2}{8a^2} \right] \int_{-\infty}^{\infty} dx \sqrt{(2\pi)^{1/2}a}\, v_0(x|\bar{z}) \\ &\quad \times \sum_{k=0}^{n} \binom{k}{n} \sqrt{(2\pi)^{1/2}ak!2^k}\, v_k(x|\bar{z}) \left(\frac{z_0-z}{\sqrt{2}a} \right)^{n-k}. \end{aligned} \quad (124)$$

By the orthonormality of v_n, only the $k = 0$ term remains non-vanishing, leading to the signal term in Eq. (40).

For the noise component,

$$\eta_n(t) = \int_{-\infty}^{\infty} dx\, \eta(x,t) v_n(x|z). \quad (125)$$

It is straightforward to check that

$$\langle \eta_n(t) \rangle = \int_{-\infty}^{\infty} dx \langle \eta(x,t) \rangle v_n(x|z) = 0, \quad (126)$$

$$\langle \eta_n(t)\eta_m(t) \rangle = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dx' \langle \eta(x,t)\eta(x',t') \rangle v_m(x|z) v_n(x'|z) = \delta_{mn}\delta(t-t'). \quad (127)$$

31

Appendix F: The Dynamics of the 2D CANN

F.1 The Eigenvalues and Eigenfunctions of $F(\mathbf{x}, \mathbf{x}'|\mathbf{z})$

Define

$$u_{m,n}(\mathbf{x}|\mathbf{z}) = u_m(x_1|z_1)u_n(x_2|z_2). \tag{128}$$

We check that $u^{m,n}$ is the right eigenfunction of $F(\mathbf{x}, \mathbf{x}'|\mathbf{z})$ and calculate the corresponding eigenvalue.

We distinguish three different cases:

- $m = n = 0$. Following the calculation in Eq. (85) and noting the integrations with respect to x_1' and x_2' can be carried out separately, we have

$$
\begin{aligned}
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} d\mathbf{x}' F(\mathbf{x}, \mathbf{x}'|\mathbf{z})u_{0,0}(\mathbf{x}'|\mathbf{z}) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} dx_1' dx_2' F(\mathbf{x}, x_1', x_2'|z_1, z_2)u_0(x_1'|z_1)u_0(x_2'|z_2) \\
&= (1 - \sqrt{1 - k/k_c})u_{0,0}(\mathbf{x}|\mathbf{z}).
\end{aligned}
\tag{129}
$$

- $m \neq 0, n = 0$ (the case for $m = 0$, $n \neq 0$ is the same). An important observation in the calculation of Eq. (84) is that for $m \geq 1$, the integral in the second term vanishes. Therefore

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} d\mathbf{x}' F(\mathbf{x}, \mathbf{x}'|\mathbf{z})u_{m,0}(\mathbf{x}'|\mathbf{z}) = \lambda_m u_{m,0}(\mathbf{x}|\mathbf{z}). \tag{130}$$

- $m \neq 0, n \neq 0$.

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} d\mathbf{x}' F(\mathbf{x}, \mathbf{x}'|\mathbf{z})u_{m,n}(\mathbf{x}'|\mathbf{z}) = \lambda_m \lambda_n u_{m,n}(\mathbf{x}|\mathbf{z}). \tag{131}$$

F.2 The network dynamics in the weak input limit

In the limit of weak inputs, we assume that the shape of the bump is unchanged and only the position of the bump varies with time, i.e., $U(\mathbf{x}, t) \approx \tilde{U}(\mathbf{x}|\mathbf{z}(t))$. We project the network dynamics in Eq. (59) on a single primary displacement mode, which is the movement of the bump along the straight line towards the stimulus location $\mathbf{z}^0$. Expressing the left-hand side of Eq. (59) in terms of the basis functions leads to

$$\tau\frac{\partial}{\partial t}\tilde{U}(\mathbf{x}|\mathbf{z}, t) = \frac{\tau U_0(2\pi)^{1/2}}{2}\left[u_{1,0}\frac{dz_1}{dt} + u_{0,1}\frac{dz_2}{dt}\right]. \tag{132}$$

On the right-hand side of Eq. (59), we note that the neuronal interaction and relaxation terms cancel each other at the steady state, leaving behind the external input. Consider first the signal

part, projected onto the components $u_{1,0}$ and $u_{0,1}$:

$$R.H.S_{\text{signal}} = \alpha U_0 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} d\mathbf{x} \exp\left[-\frac{(\mathbf{x} - \mathbf{z}^0)^2}{4a^2}\right][u_{1,0}(\mathbf{x}|\mathbf{z}) + u_{0,1}(\mathbf{x}|\mathbf{z})]. \tag{133}$$

Explicit integration yields

$$R.H.S_{\cdot\text{signal}} = -\alpha\frac{U_0(2\pi)^{1/2}}{2} \exp\left[-\frac{(\mathbf{z} - \mathbf{z}^0)^2}{8a^2}\right][(z_1 - z_1^0)u_{1,0}(\mathbf{x}|\mathbf{z}) + (z_2 - z_2^0)u_{0,1}(\mathbf{x}|vz)]. \tag{134}$$

For the noise component,

$$R.H.S_{\cdot\text{noise}} = \sigma\eta_{1,0}(t)u_{1,0}(\mathbf{x}|\mathbf{z}) + \sigma\eta_{0,1}(t)u_{0,1}(\mathbf{x}|\mathbf{z}), \tag{135}$$

where

$$\langle\eta_{1,0}(t)\rangle = \langle\eta_{0,1}(t)\rangle \quad = \quad 0, \tag{136}$$

$$\langle\eta_{1,0}(t)\eta_{1,0}(t')\rangle = \langle\eta_{0,1}(t)\eta_{0,1}(t')\rangle \quad = \quad \delta(t - t'). \tag{137}$$

Using the orthonormality of $u_{m,n}$ and expressing the result in vector form, we arrive at Eq. (72).

References

Amari, S. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics 27*, 77–87.

Ben-Yishai, R., R. Lev Bar-Or, and H. Sompolinsky (1995). Theory of orientation tuning in visual cortex. *Proc. Natl. Acad. Sci. USA 92*, 3844–3848.

Berry II, M., I. Brivanlou, T. Jordon, and M. Meister (1999). Anticipation of moving stimuli by the retina. *Nature 398*, 334–338.

Brody, C. D., R. Romo, and Kepecs (2003). Basic mechanisms for graded persistent activity: discrte attractors, continuous attractors, and dynamic representations. *Current Opinion in Neurobiology 13*, 204–211.

Camperi, M. and X.-J. Wang (1998). A model of visuospatial short-term memory in prefrontal cortex: recurrent network and cellular bistability. *J. Comput. Neurosci. 5*, 383–405.

Chow, C. and S. Coombes (2006). Existence and wandering of bumps in a spiking neural network model. *SIAM Journal of Applied Dynamical Systems 5*, 552–574.

Dayan, P. and L. Abbott (2001). *Theoretical Neuroscience.* MIT Press, Cambridge.

Deneve, S., P. E. Latham, and A. Pouget (1999). Reading population codes: a neural implementation of ideal observers. *Nature Neuroscience 2*, 740–745.

Domany, E., J. van Hemmen, and K. Schulten (1995). *Models of Neural Networks III.* Springer, Berlin.

Erlhagen, W. (2003). Internal models for visual perception. *Biol. Cybern. 88*, 409–417.

Ermentrout, B. (1998). Neural networks as spatial-temporal pattern-forming systems. *Reports on Progress in Physics 61*, 353–430.

Folias, E. and P. Bressloff (2004). Breathing pulses in an excitatory neural network. *SIAM J. Appl. Dyn. Syst. 3*, 378–407.

Fu, Y., Y. Shen, and Y. Dan (2001). Motion-induced perceptual extrapolation of blurred visual targets. *J. Neuroscience 21*, 1–5.

Funahashi, S., C. Bruce, and P. Goldman-Rakic (1989). Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *J. Neurophysiology 61*, 331–349.

Fung, C. C. A., K. Y. M. Wong, and S. Wu (2008). Dynamics of neural networks with continuous attractors. *Europhys. Lett. 84*, 18002.

Georgopoulos, A. P., J. F. Kalaska, R. Caminiti, and J. T. Massey (1982). On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. *J. Neurosci. 2*, 1527–1537.

Georgopoulos, A. P., M. Taira, and A. Lukashin (1993). Cognitive neurophysiology of the motor cortex. *Science 260*, 47–52.

Graf, A., F. Wichmann, H. Bulthoff, and B. Scholkopf (2006). Classification of faces in man and machines. *Neural Computation 18*, 143.

Griffiths, D. (2004). *Introduction to Quantum Mechanics.* Prentice Hall.

Grossberg, S. (1988). Nonlinear neural networks: principles, mechanisms and architectures. *Neural Network 1*, 17.

Gutkin, B. S., D. Pinto, and B. Ermentrout (2003). Mathematical neuroscience: from neurons to circuits to system. *J. Physiol. Paris 97*, 209–219.

Hansel, D. and H. Sompolinsky (1998). Modeling feature selectivity in local cortical circuits. In C. Koch and I. Segev (Eds.), *Methods in Neuronal Modeling: From Ions to Networks.* MIT Press, Cambridge.

Heeger, D. (1993). Modeling simple-cell direction selectivity with normalized, half-squared, linear operators. *J. Neurophysiology 70*, 1885.

Hopfield, J. J. (1984). Neurons with graded responses have collective computational properties like those of two-state neurons. *Proc. Natl. Acad. Sci. USA 81*, 3088–3092.

Jastorff, J., Z. Kourtzi, and M. Giese (2006). Learning to discriminate complex movements: biological versus artificial trajectories. *J. Vision 6*, 791.

Laing, C. R. and C. Chow (2001). Stationary bumps in networks of spiking neurons. *Neural Computation 13*, 1473–1494.

Ma, S. K. (1976). *Modern Theory of Critical Phenomena.* Benjamin, Reading, Massachusetts.

Machens, C. and C. Brody (2008). Design of continuous attractor networks with monotonic tuning using a symmetry principle. *Neural Computation 20*, 452–485.

Maunsell, J. H. R. and D. C. Van Essen (1983). Functional properties of neurons in middle

temporal visual area of the macaque monkey. i. selectivity for stimulus direction, speed, and orientation. *J. Neurophysiology 49*, 1127–1147.

Miller, P. (2006). Analysis of spike statistics in neuronal systems with continuous attractors or multiple, discrete attractor states. *Neural Computation 18*, 1268–1317.

Renart, A., P. Song, and X. Wang (2003). Robust spatial working memory through homeostatic synaptic scaling in heterogeneous cortical networks. *Neuron 38*, 473–485.

Rolls, E. T., R. G. Robertson, and P. Georges-Francois (1995). The representation of space in the primate hippocampus. *Soc Neurosci Abstr 21*, 1494.

Samsonovich, A. and B. L. McNaughton (1997). Path integration and cognitive mappping in a continuous attractor neural network model. *J. Neurosci. 7*, 5900–.

Seung, H. S. (1996). How the brain keeps the eyes still. *Proc. Acad. Sci. USA 93*, 13339–13344.

Stringer, S. M., T. P. Trappenberg, E. Rolls, and I. Aranjo (2002). Self-organizing continuous attractor networks and path integration: One-dimensional models of head direction cells. *Network: Computation in Neural Systems 13*, 217–242.

Taube, J. S. (1998). Head direction cells and the neurophysiological basis for a sense of direction. *Prog. Neurobiol. 55*, 225–256.

Trappenberg, T. (2003). Continuous attractor neural networks. In L. N. de Castro and F. J. V. Zuben (Eds.), *Recent Developments in biologically inspired computing.*

Tuckwell, H. (1988). *Introduction to theoretical neurobiology.* Cambridge University Press, Cambridge.

Wang, X. J. (2001). Synaptic reverberation underlying mnemonic persistent activitity. *Trends in Neuroscience 24*, 455–463.

Wennekers, T. (2002). Dynamic approximation of spatio-temporal receptive fields in nonlinear neural field models. *Neural Computation 14*, 1801–1825.

Wilson, H. and J. Cowan (1972). Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical Journal 12*, 1–24.

Wilson, M. A. and B. L. McNaughton (1993). Dynamics of hippocampal ensemble code for space. *Science 261*, 1055–1058.

Wu, S. and S. Amari (2005). Computing with continuous attractors: Stability and on-line aspects. *Neural Computation 17*, 2215–2239.

Wu, S., S. Amari, and H. Nakahara (2002). Population coding and decoding in a neural field: a computational study. *Neural Computation 14*, 999–1026.

Wu, S., K. Hamaguchi, and S. Amari (2008). Dynamics and computation of continuous attractors. *Neural Computation 20*, 994–1025.

Xie, X., R. Hahnloser, and S. Seung (2002). Double-ring network model of the head-direction system. *Physical Review E 66*, 041902.

Zhang, K.-C. (1996). Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory. *J. Neuroscience 16*, 2112–2126.
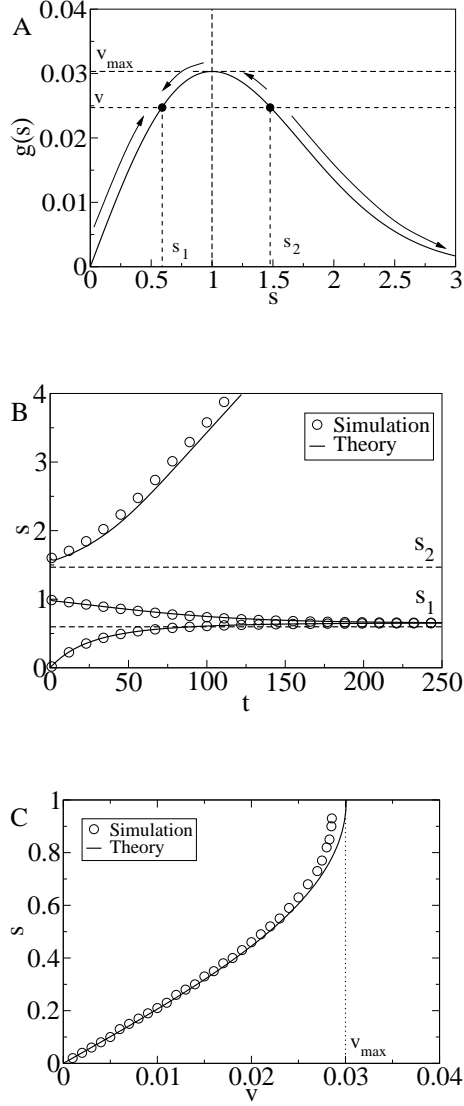
Figure 4: (A) The function $g(s)$ and the stable and unstable fixed points of Eq. (32). (B) The time dependence of the separation s starting from different initial values. $v = 0.025$. (C) The dependence of the terminal separation s on the speed v. Parameters: $N = 200$, $k = 0.5$, $a = 0.5$, $\tau = 1$, $A = \sqrt{2\pi}a$, $\rho = N/2\pi$ and $\alpha = 0.05$.
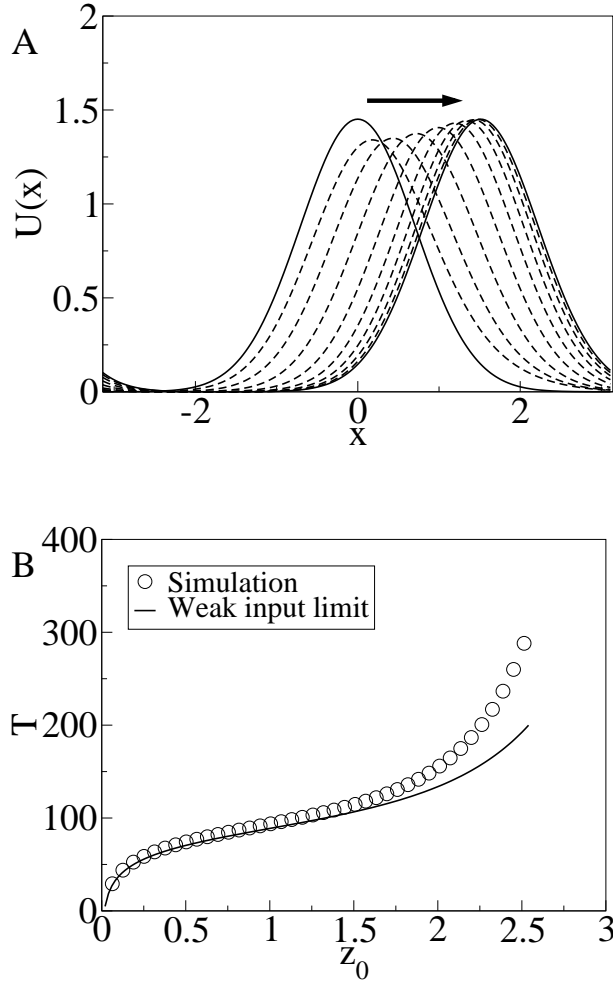
Figure 5: (A) The tracking process of the network. (B) The reaction time vs. the jump size $|z_0|$. Parameters: $\theta = \pi/200$ and the rest are the same as Fig. 4.
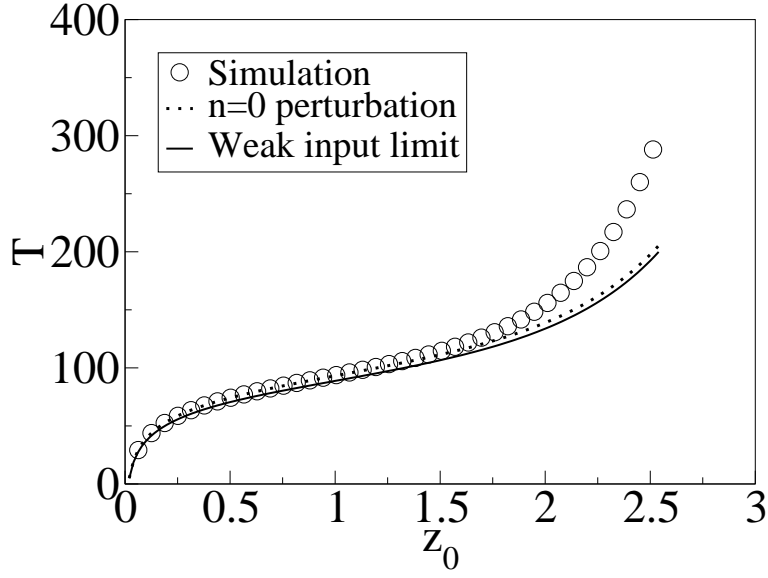
Figure 6: The reaction time vs. the jump size. Parameters: the same as Fig. 5.
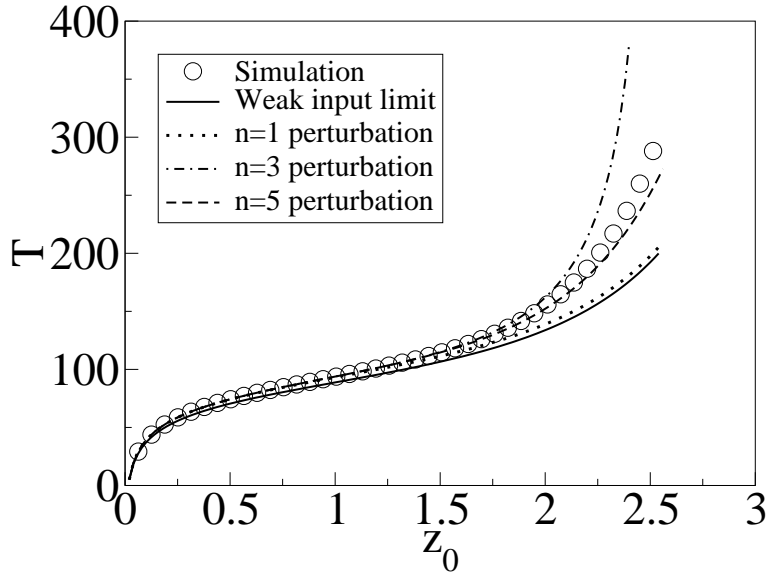
Figure 7: The reaction time vs. the jump size. The simulation result is compared with the $n = 5$ perturbation approximation. Parameters: the same as Fig. 5. Note that the predictions of the $n = 0$ and $n = 1$ perturbations are the same because of the self-consistent condition in Eq. (119). Also, the predicted reaction time given by the $n = 1$ and $n = 2$ perturbations are almost the same. So are the $n = 3$ and $n = 4$ perturbations.
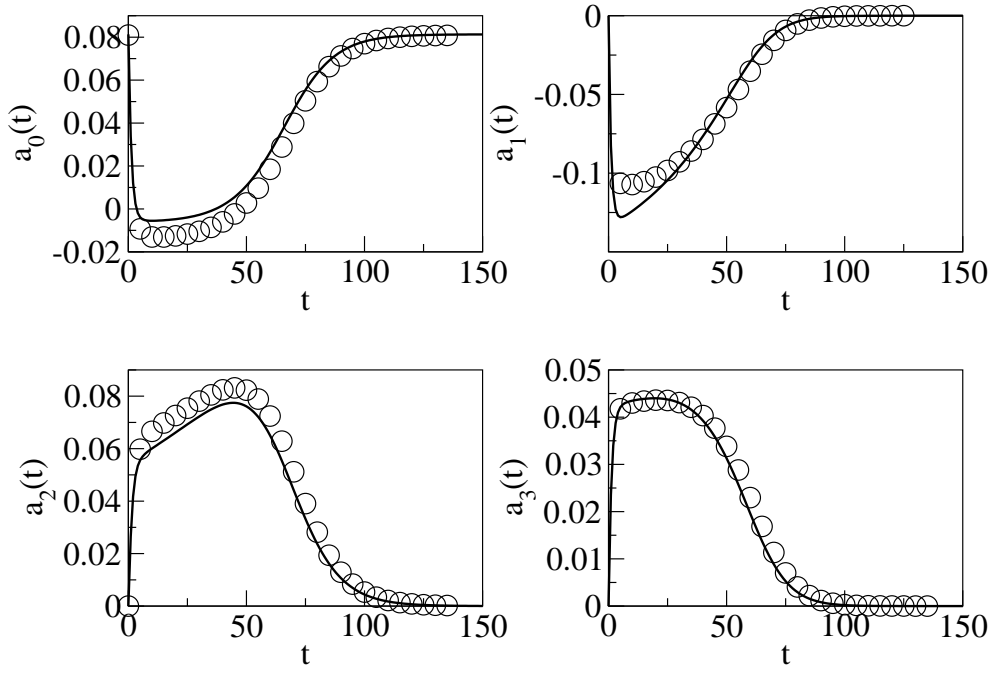
Figure 8: The first four $a_n(t)$ obtained by the $n = 10$ perturbation (dashed lines) and by projecting $U(x,t)$ onto the corresponding $v_n(x|z)$ during tracking (symbols). Parameters: $N = 200$, $\alpha = 0.05$, $a = 0.5$, $\tau = 1$, $k = 0.5$, $\rho = N/2\pi$ and $A = \sqrt{2a}\pi$.
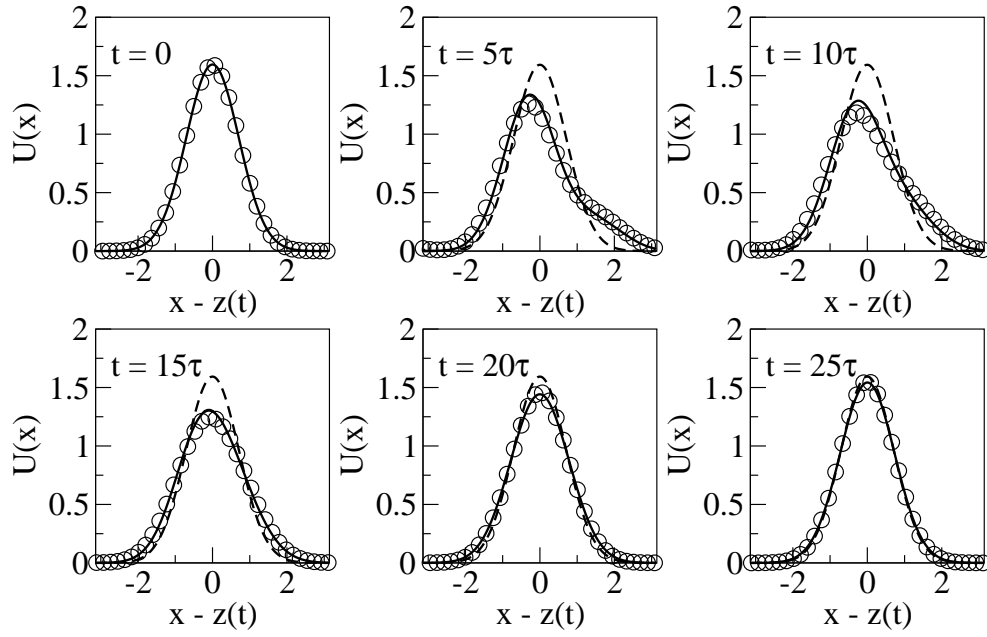
Figure 9: Snapshots of the neuronal inputs in the center-of-mass frame of the bump. Symbols: simulations. Solid lines: predicted synaptic input with the $n = 10$ perturbation. Parameters: $N = 200$, $\alpha = 0.15$, $a = 0.5$, $\tau = 1$, $k = 0.5$, $\rho = N/2\pi$ and $A = \sqrt{2a}\pi$.
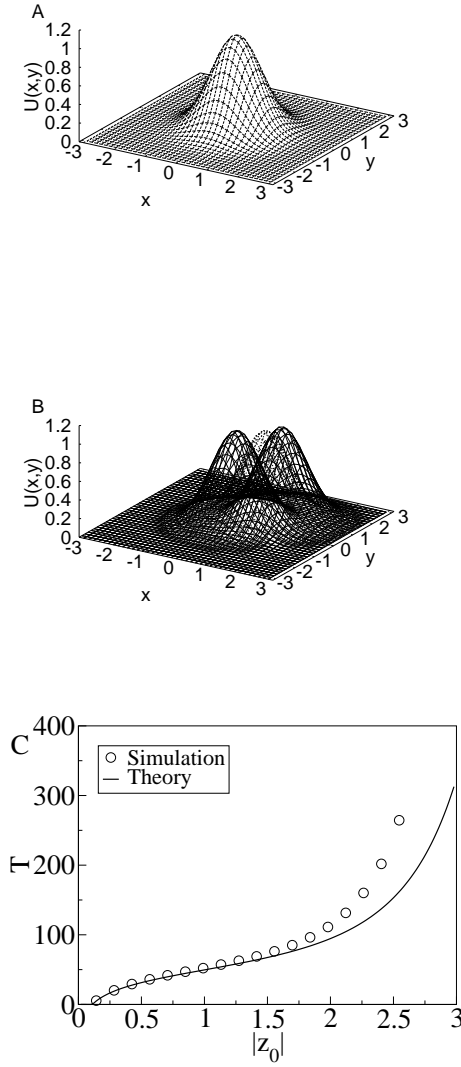
Figure 10: (A) The stationary state of the 2D CANN. (B) Illustrating the tracking process of the network; (C) The reaction time vs. the jump size. The simulation result is compared with the theoretical prediction. Parameters: $N = 40 \times 40$, $k = 0.5$, $a = 0.5$, $\tau = 1$, $A = \sqrt{2}\pi a^2$, $\rho = N/(2\pi)^2$, $\alpha = 0.05$ and $\theta = \sqrt{2}\pi/N$.