

SYDE 556/750  
Simulating Neurobiological Systems  
Lecture 14: Spatial Semantic Pointers

Chris Eliasmith

November 24 & 25, 2022



**Accompanying Readings: Dumont & Eliasmith, 2020. See [here](#).**

# Contents

<b>1 Introduction</b>	<b>1</b>
<b>2 Mathematical properties</b>	<b>1</b>
<b>3 Grid cells</b>	<b>2</b>

# 1 Introduction



**Note:** These notes are very under constructions, and mostly mathematical background. Much of the material is directly from Dumont and Eliasmith, 2020.

We define circular convolution exponentiation as:

$$B^n = \underbrace{B \otimes B \otimes \dots \otimes B}_{n \text{ times}} \quad (1)$$

This can be written:

$$B^k = \mathcal{F}^{-1}\{\mathcal{F}\{B\}^k\}, \quad k \in \mathbb{R} \quad (2)$$

where exponentiation in the fourier domain is regular exponentiation. Where:

$$\mathcal{F}\{X\} = r_x e^{i\theta_x} \quad (3)$$

Note that because we are working with exclusively unitary vectors,  $r_x = 1$ , and so the  $\theta$  phases completely determine the SSP that is being used.

We can write a 2D spatial representation as:

$$S(x, y) = X^x \otimes Y^y = \mathcal{F}^{-1}\{\mathcal{F}\{X\}^x \odot \mathcal{F}\{Y\}^y\}, \quad (4)$$

where  $\odot$  is the Hadamard (element-wise) product. In the frequency domain:

$$S(x, y) = (X^x \otimes Y^y) \quad (5)$$

$$= \mathcal{F}^{-1}(\mathcal{F}\{X\}^x \odot \mathcal{F}\{Y\}^y) \quad (6)$$

$$= \mathcal{F}^{-1}(r_x^x e^{i\theta_x x} r_y^y e^{i\theta_y y}) \quad (7)$$

$$= \mathcal{F}^{-1}(r_x^x r_y^y e^{i(\theta_x x + \theta_y y)}) \quad (8)$$

where we have left off the vector indexes for clarity, but note that the  $\theta$  for each dimension can be different.

## 2 Mathematical properties

The most critical mathematical property that SSPs have is that they preserve Euclidean relations in a much higher dimensional space. That is,

$$S(x_1, y_1) \otimes S(x_2, y_2) = S(x_1 + x_2, y_1 + y_2) \quad (9)$$

$$= X^{x_1+x_2} \otimes Y^{y_1+y_2} \quad (10)$$

So it's easy to shift a current spatial representation around without decoding the representation. We can use this to implement basic differential equations [1]:

$$S_{t+\Delta t} = (X^{\Delta x_t} \otimes Y^{\Delta y_t}) \otimes S_t, \quad (11)$$

where  $\Delta x_t$  and  $\Delta y_t$  are derived from differential equations that relate  $x$  and  $y$  to  $t$ . If the underlying dynamics are linear, we have  $\Delta x = \frac{dx}{dt} \Delta t$ . Assuming  $S_t = X^{x_t} \otimes Y^{y_t}$ , then the algebraic properties of SSPs ensure that:

$$X^{x_t} \otimes Y^{y_t} \otimes X^{\Delta x_t} \otimes Y^{\Delta y_t} \quad (12)$$

$$= X^{x_t + \Delta x_t} \otimes Y^{y_t + \Delta y_t} \quad (13)$$

$$= M_{t+\Delta t} \quad (14)$$

It's also possible to write a similar equation without discretizing time. The result of doing so gives:

$$\frac{dS}{dt} = \left( \frac{dx}{dt} \ln X + \frac{dy}{dt} \ln Y \right) \otimes S. \quad (15)$$

Of relevance to using SSPs as probability representations, it's important to note that as the dimensionality becomes sufficiently high, the expected similarity approaches:

$$X^{x_1} \cdot X^{x_2} = \text{sinc}(x_2 - x_1) \quad (16)$$

for SSPs [2].

### 3 Grid cells

Suppose ideal place cells are Gaussian bumps, and that they evenly cover a space to be represented. We would like to find the hidden layer activations  $G \in \mathbb{R}^{n_x \times n_g}$  (where  $n_g$  is the number of hidden neurons and  $n_g < n_p$ , and  $n_p$  is the number of place cells) and the matrix of read-out weights  $W \in \mathbb{R}^{n_g \times n_p}$  that minimize the reconstruction error of the place cell responses.

$$\min_{G, W} \|P - \hat{P}\|_F^2, \quad (17)$$

$$\hat{P} = GW \quad (18)$$

The optimal  $W$  for a fixed  $G$  is given by

$$W^* = (G^T G)^{-1} G^T P. \quad (19)$$

This  $W$  should be thought of as the connection weights between the final two layers of some deep neural network. The input to the full network would be low level sensory information and the output would be the place cell activity,  $P$ . The hidden layer with activations  $G$  is the last layer before the place cells, and, since  $n_g < n_p$ , it creates an information bottleneck. We are interested in finding the optimal  $G$  - a compressed representation of spatial position that is optimal for reconstructing  $P$  in a single layer.

As stated in [3], if the number of place cells is large and their receptive fields uniformly cover space (and space has periodic boundary conditions) then  $P P^T$  will approximately be a circulant matrix and its eigenvectors will be Fourier modes.

Thus, the optimal responses of hidden neurons will be linear combinations of plane waves. This will produce hidden neurons with grid-like spatial responses. Adding a non-negativity

constraint to this optimization problem will result in the activity of an individual hidden neuron being proportional to a sum of three plane waves whose wave vectors are  $120^\circ$  degrees apart. Specifically, a column of  $G$  will have entries,

$$\sum_{j=1}^3 e^{i\mathbf{k}_j \cdot \mathbf{x}_n} + e^{-i\mathbf{k}_j \cdot \mathbf{x}_n} \quad (20)$$

$$\text{where } |\mathbf{k}_j| = |\mathbf{k}_i| \quad \forall i, j \quad (21)$$

$$\sum_{j=1}^3 \mathbf{k}_j = 0 \quad (22)$$

The interference pattern of these waves will have a hexagonal grid pattern, like grid cells. Note that this is also real as the imaginary parts cancel. The first equation will show the interference patterns of the choice of  $\mathbf{k}$  vectors – i.e. be the grid cells in  $G$ .

It might be helpful to recall Euler's formula:

$$e^{ix} = \cos(x) + i\sin(x) \quad (23)$$

to show that equation 20 is:

$$\sum_{j=1}^3 e^{i\mathbf{k}_j \cdot \mathbf{x}_n} + e^{-i\mathbf{k}_j \cdot \mathbf{x}_n} \quad (24)$$

$$= \sum_{j=1}^3 \cos(\mathbf{k}_j \cdot \mathbf{x}_n) + i\sin(\mathbf{k}_j \cdot \mathbf{x}_n) + \cos(\mathbf{k}_j \cdot \mathbf{x}_n) - i\sin(\mathbf{k}_j \cdot \mathbf{x}_n) \quad (25)$$

$$= 2 \sum_{j=1}^3 \cos(\mathbf{k}_j \cdot \mathbf{x}_n) \quad (26)$$

which defines the interference pattern we see as grid-like for the appropriate choice of  $\mathbf{k}$ . Changing the orientation of the  $\mathbf{k}$  will rotate the grid, and changing the length of the  $\mathbf{k}$  will change the spatial frequency (i.e., spacing) of the grid.

## References

- [1] Aaron R. Voelker et al. "Simulating and Predicting Dynamical Systems With Spatial Semantic Pointers". In: *Neural Computation* 33.8 (July 2021), pp. 2033–2067. DOI: 10.1162/neco\_a\_01410. URL: [https://doi.org/10.1162/neco\\_a\\_01410](https://doi.org/10.1162/neco_a_01410).
- [2] Aaron R Voelker. "A short letter on the dot product between rotated Fourier transforms". In: *arXiv preprint arXiv:2007.13462* (2020).
- [3] Ben Sorscher et al. "A unified theory for the origin of grid cells through the lens of pattern formation". In: *Advances in Neural Information Processing Systems*. 2019.