EXPONENTIAL MAP MODELS AS AN INTERPRETABLE FRAMEWORK FOR GENERATING NEURAL SPATIAL REPRESENTATIONS

Anonymous authors
Paper under double-blind review

ABSTRACT

A fundamental challenge in neuroscience and AI is understanding how physical space is mapped into neural representations. While artificial neural networks can generate brain-like spatial representations, such as place and grid cells, their "black-box" nature makes it difficult to determine if these representations arise as general solutions or as artifacts of a chosen architecture, objective function, or training protocol. Critically, these models offer no guarantee that learned solutions for core navigational tasks, like path integration (updating position from self-motion), will generalize beyond their training data. To address these challenges, we introduce a first-principles framework based on an exponential map model. Instead of using deep networks or gradient optimization, the presented model uses generator matrices to map physical locations into neural representations through the matrix exponential, creating a transparent framework that allows us to identify several exact algebraic conditions underlying key properties of neural maps. We show that path invariance (ensuring location representations are independent of traversal route) is achieved if the generators commute, while translational invariance (maintaining consistent spatial relationships across locations) demands generators producing orthogonal transformations. We also show that preserving the metric of flat space requires the eigenvalues of the generator matrices to form sets of roots of unity. Finally, we demonstrate that the proposed framework constructs diverse biologically relevant spatial tuning, including place cells, grid cells, and context-dependent remapping. The framework we propose thus offers a transparent, theoretically-grounded alternative to "black-box" models, revealing the exact conditions required for a coherent neural map of space.

1 INTRODUCTION

A fundamental challenge in neuroscience and artificial intelligence is to understand the mapping from physical space to the representational space of neural population activity. In the mammalian brain, such representations are strongly associated with the hippocampal formation, which contains specialized neurons that encode spatial information. Most famously, place cells (O'Keefe & Dostrovsky, 1971) fire within specific, localized areas of an environment known as place fields, while grid cells (Hafting et al., 2005) fire in a periodic hexagonal pattern that tessellates the environment and is believed to provide a neural metric for space (Ginosar et al., 2023). Together, these and other spatially-tuned cells form a rich, high-dimensional representation of an animal's location. This neuronal spatial map abruptly reorganizes in response to environmental changes, a phenomenon known as remapping, indicating that neurons also encode the environment's identity (Leutgeb et al., 2004; Fyhn et al., 2007). While the firing patterns of these spatial neurons are well-characterized, the principles governing their emergence remain unclear.

In recent years, deep learning models, particularly recurrent neural networks (RNNs) trained to solve navigation tasks, have been shown to *learn* representations that resemble biological place and grid cells (Banino et al., 2018; Cueva & Wei, 2018; Sorscher et al., 2022; Whittington et al., 2020). These findings are significant, as they strongly suggest that spatial tuning is a normative solution to the demands of navigation. However, the "black-box" nature of deep neural networks makes it difficult to disentangle whether their learned representations reflect fundamental principles of navigation or

are artifacts of a chosen architecture, objective function, or training protocol (see Fig. 1a) for an illustration). Another key limitation of this approach is that deep learning models offer no guarantee that their learned solutions will generalize beyond the training data. In contrast, animals are able to seamlessly navigate vast, novel environments. To understand how biological brains solve these tasks so readily, there is a need for models that allow for exact and interpretable solutions to navigation problems.

In this work, we construct spatial representations using an exponential map model. Instead of relying on neural networks or gradient-based optimization, the presented approach builds representations from a transparent mathematical foundation. The core component of the model is a set of generator matrices that directly map a spatial location to a neural population firing rate vector. This construction allows us to derive the exact algebraic conditions required for a coherent neural spatial map. For a neural spatial map to be useful, it must support core navigational computations. One of the most fundamental is *path integration*, the process by which a navigator estimates its position by integrating self-motion cues. This process introduces a critical self-consistency problem: For the map to be coherent, the representation of a location must be independent of the path taken to reach it. We show that path-independent representations required for reliable path integration are guaranteed if the model's generator matrices commute. Furthermore, we find that equinorm representations, previously used as a learning constraint in neural networks (Schaeffer et al., 2023; Xu et al., 2022), arise naturally from generators that produce translationally invariant similarity structures—a desirable property for navigation in open-field environments. We also show that preserving the metric of flat space (Xu et al., 2022) requires the eigenvalues of the generator matrices to form sets of roots of unity on discrete rings in frequency space. When all of these properties are taken into account, the generated spatial representations are similar to those spatial cells in the brain, depending on a choice of symmetry. Finally, we demonstrate that this framework can be seamlessly generalized from preserving the metric of space to preserving the similarity of more general inputs, which we use to model remapping. A conceptual overview of the proposed framework and the key spatial map properties we address are presented in Fig. 1.

Despite its simplicity, the proposed framework is powerful enough to construct a wide variety of biologically plausible tuning curves, including place cells, grid cells, and context-dependent remapping, from the same underlying mechanism. By grounding spatial representations in a clear algebraic structure, the presented work provides a theoretically-grounded alternative to black-box models, revealing exact and interpretable principles that underpin a coherent neural map of space.

2 Results & Discussion

2.1 Constructing Spatial Representations with an Exponential Map

A spatial representation is, in broad terms, a map that assigns a neural population vector to every spatial location, as exemplified in Fig. 1b). For a 2D space with Cartesian coordinates (x, y), the representation at a point is a population vector $\mathbf{p}(x, y) \in \mathbb{R}^N$. Each of the N components of this vector can be thought of as the firing rate of a neuron, making the vector a point in an N-dimensional state space that captures the activity of the entire neural ensemble. We build upon previous modeling approaches Gao et al. (2021); McNamee et al. (2021); Xu et al. (2022) and define this map using the matrix exponential:

$$\mathbf{p}(x, y) = e^{xG_x + yG_y}\mathbf{p}_0, \tag{1}$$

where $G_x, G_y \in \mathbb{R}^{N\times N}$ are *generator matrices* for the cardinal directions and $\mathbf{p}_0 = \mathbf{p}(x_0, y_0)$ is the representation at some origin point. Intuitively, the generator matrices define how locations in physical space translate into transformations in the high-dimensional neural state space. The exponential map then composes these transformations to "transport" the origin vector, $\mathbf{p}_0$ to a population vector at any target location (x, y).

Equation (1) does what we intended it to; for each location (x, y), it assigns a population vector, and provides a constructive method for generating a spatial map. However, without further constraints, an arbitrary choice of generators could produce a map that is ill-suited for navigation. For instance, the representation could end up being trivial (all locations map to the same vector) or ambiguous (multiple locations map to the same vector). As we will show, the power of this framework lies

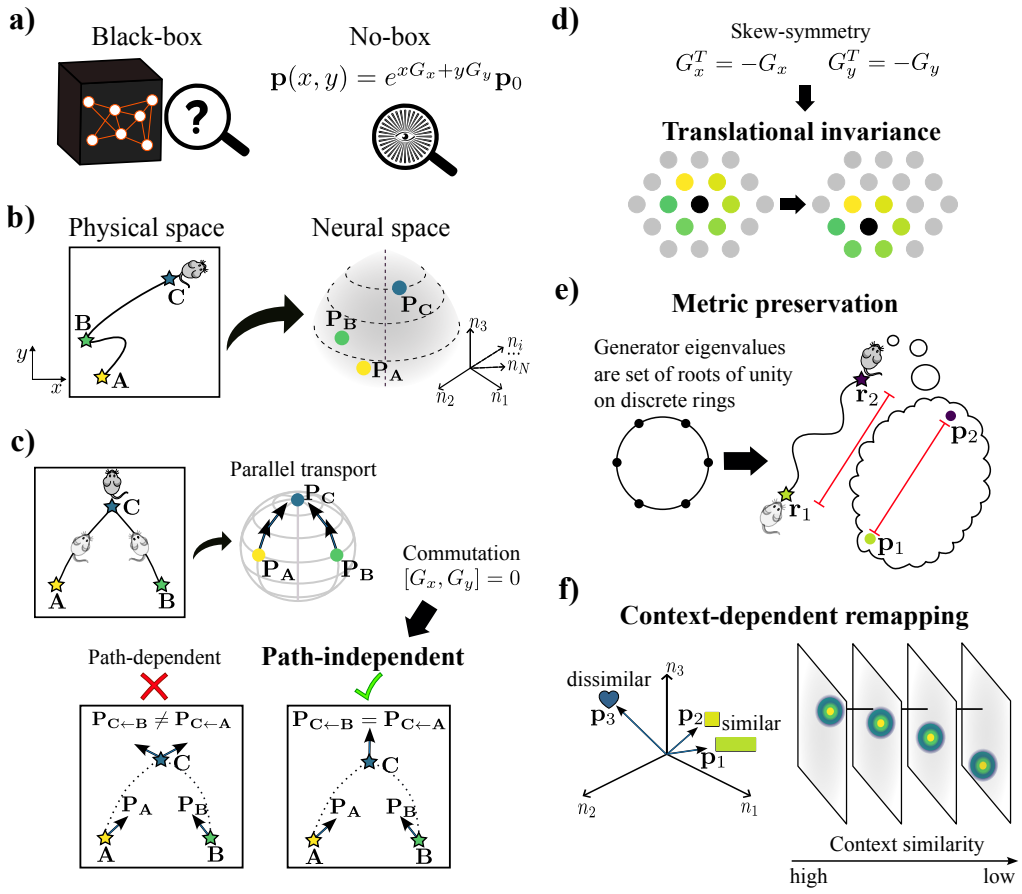

Figure 1: Conceptual overview of the proposed framework and results. **a)** Deep learning models are "black-boxes" that learn spatial representations, but the underlying principles are obscured by the complexities of architecture, training, and objective functions. The exponential map model is a transparent "no-box" alternative, using generator matrices (G_x, G_y) to construct a representation $\mathbf{p}(x, y)$. **b)** A neural population vector, which captures the activity of the entire neural ensemble, is assigned to every location, mapping physical space to a neural representational space. **c) Path Invariance:** Path integration can be viewed as a form of parallel transport, where a vector representing the neural representation is moved along a trajectory in a high-dimensional state space. Traversing a curved manifold can induce a net transformation in the vector at a point that is dependent on the traversal route. By imposing simple, interpretable algebraic constraints on the model's generators, we can directly enforce fundamental properties. Path invariance is guaranteed if the generators commute. **d) Translational Invariance:** Making the generator matrices skew-symmetric $(G^T = -G)$ imposes several biologically-relevant properties on the representation. First, it ensures that spatial relationships are consistently maintained across locations (translational invariance). Second, skew-symmetric generators produce orthogonal representations, meaning the population vector $\mathbf{p}(x, y)$ maintains a constant norm across the entire space. **e) Metric Preservation:** Preserving the geometry of flat space requires the generator eigenvalues to form sets of roots of unity on discrete rings in frequency space, which, for certain symmetry orders gives rise to grid-like patterns. **f) Remapping:** Generalizing the framework to non-spatial inputs, like a context signal, allows the model to produce distinct spatial maps for different contexts, mimicking remapping.

in its transparency, allowing us to derive precise algebraic conditions on G_x and G_y that guarantee properties essential to navigation.

2.2 From Spatial Representation to Path Integration and path Independence

Path integration is a crucial skill possessed by most animals, wherein one's location is inferred by integrating past location and self-motion information. In terms of the representation in Eq. (1), path integration is realized if

$$\mathbf{p}(x + \Delta x, y + \Delta y) = Q(\Delta x, \Delta y)\mathbf{p}(x, y). \tag{2}$$

Intuitively, we can say that we can perform path integration, if, for any past location (x, y) and the corresponding population vector $\mathbf{p}(x, y)$ we can arrive at the correct population vector $\mathbf{p}(x + \Delta x, y + \Delta y)$ at the new location $(x + \Delta x, y + \Delta y)$ through some operation Q that only depends on the displacement. Inserting our spatial representation from Eq. (1), we find that we want

$$e^{(x+\Delta x)G_x+(y+\Delta y)G_y}\mathbf{p}_0 = Q(\Delta x, \Delta y)e^{xG_x+yG_y}\mathbf{p}_0.$$

This equality suggests that we want

$$Q(\Delta x, \Delta y) = e^{\Delta x G_x+\Delta y G_y}.$$

However, the exponential function in Eq. (1) is a matrix exponential, which behaves differently from the regular exponential function. In particular, the Baker-Campbell-Hausdorff formula dictates that

$$e^A e^B = e^{A+B+\frac{1}{2}[A,B]+\frac{1}{12}[A,[A,B]]+\frac{1}{12}[B,[B,A]]+\ \ldots},$$

where $[A, B] = AB - BA$ is the commutator between matrices A and B. However, this immediately reveals that if the generator matrices G_x, G_y commute, $[G_x, G_y] = 0$, then Eq. (2) is automatically satisfied for any displacement, as

$$[aG_x + bG_y, a'G_x + b'G_y] = (ab' - a'b)[G_x, G_y],$$

for all (a, b) and (a', b'). Thus, if the generators commute, the model can path integrate exactly and indefinitely! An important effect of this choice is that the representation is path-invariant (as illustrated in Fig. 1c), meaning that the population vector at a point does not depend on the path taken to it. This is also demonstrated explicitly in Appendix B. Going forward, we therefore demand that G_x, G_y commute, which ensures that the representation $\mathbf{p}$ is path-integration compatible, as enacted by Eq. (2). Next, we demonstrate that commuting generator matrices enable an explicit construction that allows us to specify the similarity structure of the spatial representation.

2.3 Orthogonal Transformations for Egocentric Navigation

Equipped with a path integration-compatible model, we can begin to consider what makes for a good or useful representation. However, before designing a representation, we need to know how to compare representations at different locations. We hold that this is most easily encoded in the similarity structure of the representation, that is, the similarity between population vectors at different locations. Considering the path-integration compatible model Eq. (2), we are therefore interested in the quantity

$$C(x, y, \Delta x, \Delta y) = \frac{\mathbf{p}(x, y)^T Q(\Delta x, \Delta y)\mathbf{p}(x, y)}{|\mathbf{p}(x, y)||Q(\Delta x, \Delta y)\mathbf{p}(x, y)|},$$

which is just the cosine similarity between the population vector at a location (x, y) and a population vector at some other location $(x + \Delta x, y + \Delta y)$ arrived at through path integration. However, using that $(e^A)^T = e^{A^T}$, and again demanding commutativity of all involved matrices, the similarity becomes

$$C(x, y, \Delta x, \Delta y) = \mathbf{p}_0^T e^{x(G_x^T+G_x)+y(G_y^T+G_y)+\Delta x G_x+\Delta y G_y}\mathbf{p}_0/Z, \tag{3}$$

where $Z(x, y, \Delta x, \Delta y)$ is shorthand for the norm factor in the original similarity expression. Surprisingly, Eq. (3) reveals that similarities are position-independent, or equivalently, translation invariant, if the generator matrices G_x and G_y are skew-symmetric, because the exponents cancel, as illustrated in Fig. 1d). When G_x and G_y are both skew-symmetric, linear combinations of the two are also skew-symmetric. For a skew-symmetric matrix A, the corresponding matrix exponential, e^A, is orthogonal. For an orthogonal exponential map, the representation generated by Eq. (1) is guaranteed to be of constant norm and so $Z = |\mathbf{p}_0|^2$. Going forward, we will demand that

$$G_x^T = -G_x \quad G_y^T = -G_y,$$

which ensures that

$$C(\Delta x, \Delta y) = \frac{\mathbf{p}_0^T e^{\Delta x G_x + \Delta y G_y} \mathbf{p}_0}{|\mathbf{p}_0|^2}, \tag{4}$$

meaning the similarity only depends on the displacement. From a navigational perspective this is important: For one, it is a sensible choice in the open-field regime, where no locations are inherently special, meaning that there is no reason for similarities to appear different at particular locations. Second, it allows the model to make spatial inferences (such as computing distances; see Section 2.4) without absolute positional information. Thus, a representation generated by orthogonal transformations can make for an ideal basis for egocentric navigation, for example, in novel environments.

We also note that recent models of spatial cells have included norm constraints which have been shown to be conducive to grid-like representations Gao et al. (2021); Dorrell et al. (2023); Xu et al. (2022); Schaeffer et al. (2023). However, similarity translational invariance has not, to the best of our knowledge, been explored explicitly in the past in the context of spatial representations. Furthermore, similarity invariance could be interesting to study also in other task domains. As an example, batch and in particular layer normalization Ba et al. (2016); Ioffe & Szegedy (2015) are reminiscent of norm constraints, and can greatly improve learning performance in neural networks. Investigating whether this could be facilitated by some between-representation similarity invariance could provide valuable insights into the goings-on of deep neural networks.

Once G_x and G_y are skew-symmetric, they may each be expressed in a very useful block diagonal form with

$$G_x = R^T \Sigma_x R \quad \text{and} \quad G_y = R^T \Sigma_y R, \tag{5}$$

where $R \in \mathbb{R}^{N \times N}$ is an orthogonal matrix, and Σ_x and Σ_y are block diagonal, with 2×2 skew symmetric blocks along the diagonal. Note that for simplicity, we will restrict ourselves to the case where N is even. In this case, the non-zero entries of Σ_x and Σ_y are the imaginary parts of the eigenvalues of G_x and G_y, which come in conjugate pairs $\pm(i\lambda_{i,x}, i\lambda_{i,y})$. Notice that this choice ensures that G_x and G_y commute, as the 2×2 skew symmetric matrices that make up the blocks of Σ_x and Σ_y commute, and $R^T R = R R^T = I$. With these prerequisites, the similarity admits the particularly simple form

$$C(\Delta x, \Delta y) = \sum_i^N \alpha_{0,i}^2 \cos(\lambda_{i,x} \Delta x + \lambda_{i,y} \Delta y), \tag{6}$$

where $\boldsymbol{\alpha}_0 = R \frac{\mathbf{p}_0}{|\mathbf{p}_0|}$ and $\lambda_{i,x}, \lambda_{i,y}$ being the imaginary part of the i-th eigenvalues of G_x and G_y, respectively (see Appendix E for a derivation).

The translational invariance induced by skew-symmetric generators comes with a non-trivial advantage: The similarity is invariant to a constant, non-spatial shift, similar to remapping behavior (see Section 2.5 for details and Fig. 1f) for an illustration) (Leutgeb et al., 2004; Fyhn et al., 2007). Between-context similarities share the same similarity function as the spatial case, except that translations are taken between contexts, not locations. Notably, this enables a single, static model to produce different spatial representations when comparing across contexts.

2.4 Preserving the metric of flat space

Given a spatial representation and a notion of representational similarity, we can finally consider what properties the representation should possess. As proposed by (Gao et al., 2021; Xu et al., 2022), we champion that one of the foundational properties of any spatial representation is its translation of physical distances into distances on a neural manifold. More specifically, we restrict ourselves to the open field (where all directions and locations are, for all purposes, equal). In this case, one would not expect distances to appear warped in any particular location or direction, and thus the metric induced by Eq. (1) should match the flat metric, at least up to a constant factor (a so-called conformal isometry (Xu et al., 2022)).

When we impose this requirement on the path integrating, orthogonal representation, we arrive at a simple condition on the eigenvalues of the generators G_x and G_y: If these form sets of roots of unity, then the representation preserves the flat metric (see Appendix D for details and Fig. 1e) for an illustration). Concretely, we write $\lambda_{i,x} = k_i \cos(\phi_i)$ and $\lambda_{i,y} = k_i \sin(\phi_i)$ in polar coordinates

(which are, again, the imaginary parts of the ith eigenvalues of G_x and G_y), a flat metric-preserving representation satisfies

$$\sum_{j}^{N/2} \rho_j^2 e^{2i\phi_j} = 0,$$

with $\rho_i = \alpha_i k_i$ being shared by conjugate eigenvalues (see Appendix D for specifics). Said differently, if the eigenvalue angles ϕ_i are evenly spaced on discrete rings, the representation preserves the flat metric. More precisely, for a given ring of radius ρ_m, there are M eigenvalues which are evenly spaced on the ring, which may possess some shared orientation offset φ_i

$$\phi_i \to \varphi_i + \pi \frac{i}{M},$$

with $i = 0, 1, ..., M-1$. To see what kind of representation a particular symmetry M produces, we first note that for orthogonal matrices, the entries of the representation $\mathbf{p}$ can be viewed as mixtures of 2D plane waves, which we denote $\boldsymbol{\alpha}$ (see Appendix C) . However, the exact mixture is determined by the choice of matrix R, as $\mathbf{p}_0 = R^T\boldsymbol{\alpha}$. Figure 2 shows example representations $\mathbf{p}$ and plane waves $\boldsymbol{\alpha}$ for different symmetries M, when R is a randomly sampled orthogonal matrix (see Appendix A for details). Also shown is the similarity function relative to the origin.

Considering the case of a single ring, we see that lower order symmetries such as $M = 2$ and 3, produces mixtures of plane waves oriented at 90 and 60 degrees, respectively. Notably, this results in grid-like representations, with square-type grids for $M = 2$, and hexagonal-type grids for $M = 3$. Notice, however that some representations are not purely grid-like, due to the random mixing by R. For greater values of M, however, the representation becomes heterogenous, and without any obvious periodicity. While the representation is strongly influenced by a choice of R, the similarity is independent of it (it only depends on $\boldsymbol{\alpha}$). Markedly, with increasing M, the similarity becomes approximately radial, and for $M = 20$, the similarity function is an approximate Bessel function, as predicted in Appendix F for a single ring of eigenvalues.

Besides uncovering a general condition for metric preservation, we also find that the admissible solutions allow for the modular organization found in grid cells in the brain (Stensola et al., 2012). Furthermore, when considering the similarity function (see Appendix F), we find that if modules with the same spacing form roots of unity in their orientation, and the symmetry of the module orientation is co-prime to that of the pattern, the representation is head-direction independent over a large spatial range. In the same vein, we also find that if the relative spacing of different modules is proportional to the zeroes of the Bessel function J_0, then the similarity function approximates a Fourier-Bessel series, which could be used to construct a range of different radially symmetric similarity functions, which could conceivably be used tune navigation to certain length scales. Furthermore, we find that the average ratio of successive Bessel function zeros used to construct a Fourier-Bessel series falls in the same variability range as the experimentally observed ratio of $\sqrt{2}$ between grid cell module spacings (Stensola et al., 2012), suggesting a possible link between our findings and the organization of entorhinal grid cells (see Appendix F).

2.5 Similarity preservation and remapping

With our choice of generators in Eq. (5) and a choice of similarity function, we are able to generate spatial representations for a single environment up to a choice of orthogonal transformation R. However, animals are capable of distinctly encoding a variety of both spatial and non-spatial (such as room smell or identity) information through so-called remapping. In this section, we will demonstrate that we can effectively model remapping behavior, extending our model to a much larger class of representations.

This result follows from noting that the between-representation similarities only depend on the *spatial* displacement between them. If, on the other hand, the representation in Eq. (1) were to encode non-spatial information and we fix spatial locations, representational similarities depend only on the change in the non-spatial input. To see this, we can consider the encoding of a simple signal; a global scalar signal s, such as the smell of the recording environment. We then encode this in the exact same manner as spatial coordinates, by defining

$$\mathbf{p}(x, y, s) = e^{xG_x + yG_y + sG_s}\mathbf{p}_0. \tag{7}$$

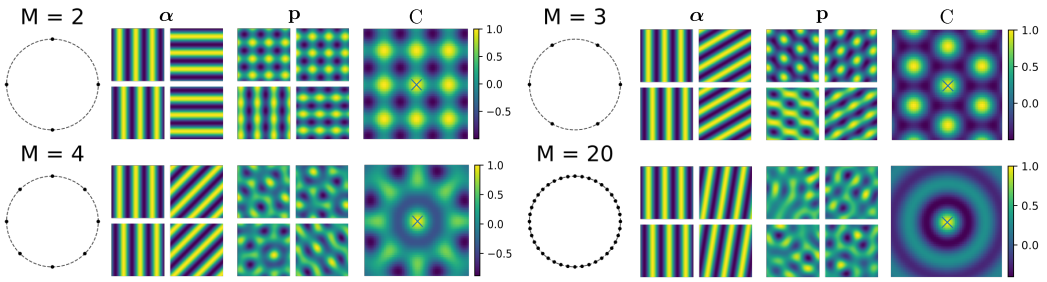

Figure 2: Example plane waves ($\boldsymbol{\alpha}$) and corresponding representations ($\mathbf{p}$) alongside the similarity function (C) relative to the origin (black cross) for representations whose eigenvalues form sets of roots of unity, with symmetries M on a single ring. For each eigenvalue (imaginary part indicated by black dot), the corresponding conjugate eigenvalue is also shown. Representations were formed using generators with a single set root-of-unity solution with varying M, a random orthogonal matrix R, and $\mathbf{p}_0 = R^T\mathbf{1}$.

The representation is now coupled to the non-spatial signal by a generator matrix G_s. Notably, G_s can be made to inherit the favorable properties of the spatial representation, by setting $G_s = R^T\Sigma_s R$, as with the spatial case. We can then consider the similarity between representations, for two distinct context signals s and s' while keeping spatial location fixed:

$$\mathbf{p}(x, y, s)^T\mathbf{p}(x, y, s') = \mathbf{p}^T e^{(s-s')G_R}\mathbf{p} = \mathbf{p}^T e^{\Delta s G_R}\mathbf{p},$$

which inherits the expression for the similarity from the spatial case. From this we can conclude that when comparing between different contexts, representations can change even as spatial location remains fixed (the nature of the modulation is codified by the choice of G_s), similar to remapping in spatial cells (Leutgeb et al., 2004; Fyhn et al., 2007).

Encoding non-metric information such as a context signal raises an additional challenge compared to the purely spatial case, as there is no clear metric or distance function that should be preserved. Instead, we can consider the more general case, where G_s should be chosen so that similar context signals produce similar representations, and dissimilar contexts result in dissimilar representations. This kind of input similarity preservation has been studied previously, and has been shown to result in localized receptive fields similar to place fields, when applied to spatial similarity (Sengupta et al., 2018; Pettersen et al., 2024). So, how could we choose G_s to perform similarity preservation? Returning to the similarity function, we note that it may be written as

$$C(\Delta s) = \sum_i \alpha_{0,i}^2 \cos(\Delta s \lambda_{i,s}), \tag{8}$$

where $\lambda_{i,s}$ denotes the imaginary part of the ith eigenvalue of G_s, whenever spatial location is fixed. Since this expression is derived from the cosine similarity, it is bounded by $[-1, 1]$. As the goal is similarity preservation, we want for $C(\Delta s)$ to approximate a function that decays with increasing Δs to some baseline level at which inputs are deemed dissimilar. We can approximate several such functions, by noting that Eq. (8) is a cosine series with non-negative coefficients. In fact, it may be viewed as a discrete approximation of the inverse Fourier transform of a symmetric function with a non-negative Fourier spectrum, as

$$\begin{aligned} f(x) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} F(k)e^{ikx}dk = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} F(k)\cos(kx)dk \\ &\approx \frac{1}{\sqrt{2\pi}N}\sum_i^N \frac{F(k_i)}{p(k_i)}\cos(k_i x), \end{aligned}$$

where the approximation is a Monte Carlo estimate of the integral using importance sampling, with k_i sampled according to some density $p(k_i)$. While there are several functions that meet the specified criteria, an especially important example is the Gaussian function, whose Fourier transform is itself a Gaussian (which is symmetric and non-negative):

$$\text{if} \quad f(x) = e^{-\sigma^2 x^2}, \quad \text{then} \quad F(k) = \frac{1}{\sqrt{2\sigma^2}}e^{-k^2/(4\sigma^2)},$$

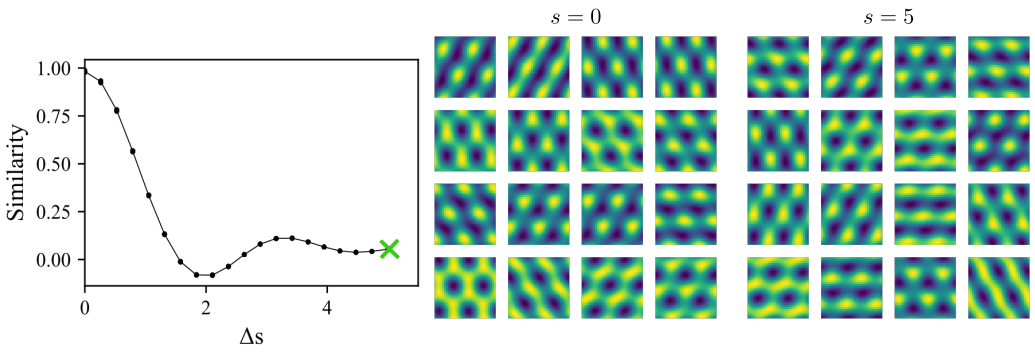

Figure 3: Between-context similarity as a function of context separation. Also inset are example ratemaps for two distant context values ($s = 0$, $s = 5$) corresponding to $\Delta s = 5$. Representations were formed using generators with 10 identical root-of-unity solutions with $M = 3$, a random orthogonal matrix R, and $\mathbf{p}_0 = R^T\mathbf{1}$.

so if we sample the eigenvalues $\lambda_{i,s}$ from a Gaussian distribution with density

$$p_i(\lambda_{i,s}) = \frac{1}{\sqrt{4\pi\sigma^2}} e^{-\lambda_{i,s}^2/(4\sigma^2)},$$

the series coefficients simplify to

$$\alpha_{0,i}^2 = \frac{1}{\sqrt{2\pi}N} \frac{F(\lambda_{i,s})}{p_i(\lambda_{i,s})} = \frac{1}{N},$$

which ensures that

$$C(\Delta s) \approx e^{-\sigma^2 \Delta s^2},$$

meaning similar contexts are highly similar, while dissimilar contexts become decorrelated, as desired. Going forward, we will set $\sigma = 1$. Notice that this result also requires us to choose $\alpha_i = 1/\sqrt{N}$, which is easily achieved if $R\mathbf{p}_0 = \frac{1}{\sqrt{N}}\mathbf{1}$, or $\mathbf{p}_0 = \frac{1}{\sqrt{N}}R^T\mathbf{1}$.

To demonstrate this remapping behavior in action, we took a metric-preserving spatial representation, consisting of 10 sets of identical root-of-unity solutions, and extended it to encode a non-spatial signal s, according to Eq. (7). Note that we include multiple sets of roots of unity, as the Monte Carlo estimate requires a larger number of terms (that is, cells) to provide a fair approximation of the Gaussian. The eigenvalues of the generator G_s were sampled according to a normal distribution, as described before. The resulting between-context similarity is shown in Fig. 3, for different contextual displacements. Notably, similarities decay with increasing context dissimilarity. Also shown are example rate maps of unit activity, which demonstrate that spatial representations shift between contexts, mimicking remapping behavior (Fyhn et al., 2007). Note that spatial similarities are preserved as long as the context signal is fixed, meaning that for a particular s, the spatial similarity is as shown in Fig. 2 for $M = 3$.

Lastly, we find that if the metric preservation requirement is relaxed, and we instead demand only similarity preservation in space using the same Gaussian similarity function for spatial locations (with eigenvalues of the generator matrices G_x, G_y sampled from a normal distribution, following the Fourier approach described previously), the resulting spatial similarity is approximately Gaussian (see Appendix G). In this case, spatial representations become heterogeneous and more strongly tuned to specific locations, resembling the tuning curves of place cells (O'Keefe & Dostrovsky, 1971). Thus, by altering the similarity function, the exponential map model can generate a diverse range of spatial tuning curves observed in the brain.

3 Conclusion

In this work, we introduced a first-principles framework for generating neural spatial representations using an exponential map model. By leveraging generator matrices and the matrix exponential, we bypassed the "black-box" nature of deep learning models, allowing for a transparent and

theoretically-grounded investigation into the principles of neural navigation. We derived the exact algebraic conditions required for a coherent map of space. Specifically, we demonstrated that commuting generators guarantee path-independent representations, a critical component for accurate path integration. Furthermore, we showed that by constraining the generators to be skew-symmetric, and thus producing orthogonal transformations, the resulting representations exhibit translational invariance in their similarity structure, an ideal property for egocentric navigation in open-field environments. We also established that preserving the flat metric of Euclidean space requires the generator eigenvalues to form sets of roots of unity on discrete rings in frequency space. Despite its mathematical simplicity, the proposed framework is capable of constructing a diverse range of biologically plausible spatial tuning, including grid cells and place cells, and modeling context-dependent remapping by extending the same principles to non-spatial inputs. This work offers an interpretable alternative to conventional deep learning approaches, revealing the fundamental mathematical structures that may underpin how the brain represents and navigates through space.

4 Limitations and future work

While our framework provides a transparent account of how coherent spatial maps can be formed, it has several limitations that open avenues for future research. The current model is primarily developed for navigation in flat, open-field environments. Animals, however, must navigate complex, curved, and obstacle-laden spaces. Future work should explore how the generator framework can be extended to represent non-Euclidean geometries, potentially by introducing position-dependent or non-commuting generators that reflect the local topology and geometry of the environment.

Second, our remapping model considers only a simple scalar context signal. A natural next step is to generalize this to handle high-dimensional, structured inputs, such as visual scenes or complex sensory cues, to model how environmental identity and spatial location are integrated into a unified representation. This would connect our algebraic approach more closely with the rich, multi-modal inputs that biological systems and artificial agents must process.

Third, while most model parameters have been fixed by simple geometric considerations, there is still a matter of finding conditions that fix the choice of the orthogonal matrix R. In this work, we have only considered randomly sampled orthogonal matrices, and it is evident that this choice strongly influences the appearance of the generated representations by mixing the underlying plane waves. Interestingly, however, this freedom can be dissociated from the representational similarity structure. As shown when modeling similarity preservation for remapping, by selecting an appropriate initial vector $\mathbf{p}_0$ the similarity function C becomes independent of the specific choice of R. This suggests that while individual tuning curves are shaped by R, the overall geometry of the neural map need not be. Future work should explore if meaningful energy constraints (Cueva & Wei, 2018) or non-negativity (Sorscher et al., 2022) constraints could mandate particular matrices R. This could, in turn, drive generated representations to be even more closely related to the striking hexagonal or sparse place-bound tunings observed in the brain.

Finally, while we propose exact conditions for properties like path integration and metric preservation, we do not specify the biological mechanisms or learning rules that would allow a neural circuit to satisfy these constraints. The proposed framework is descriptive, not prescriptive, in how these solutions are achieved. Investigating how biologically plausible learning rules, such as Hebbian plasticity or gradient-based learning in recurrent neural networks, might converge to these mathematically ideal solutions is a critical direction for future inquiry. For instance, could the norm and similarity constraints explored here serve as powerful priors or regularizers for training more robust and generalizable navigation agents? Answering such questions will help bridge the gap between our theoretical work and its implementation in both biological and artificial neural systems.

References

Michael I. Anderson and Kathryn J. Jeffery. Heterogeneous Modulation of Place Cell Firing by Changes in Context. *The Journal of Neuroscience*, 23(26):8827–8835, October 2003. ISSN 0270-6474, 1529-2401. doi: 10.1523/JNEUROSCI.23-26-08827.2003. URL https://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.23-26-08827.2003.

Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E. Hinton. Layer normalization, 2016. URL https://arxiv.org/abs/1607.06450.

Andrea Banino, Caswell Barry, Benigno Uria, Charles Blundell, Timothy Lillicrap, Piotr Mirowski, Alexander Pritzel, Martin J. Chadwick, Thomas Degris, Joseph Modayil, Greg Wayne, Hubert Soyer, Fabio Viola, Brian Zhang, Ross Goroshin, Neil Rabinowitz, Razvan Pascanu, Charlie Beattie, Stig Petersen, Amir Sadik, Stephen Gaffney, Helen King, Koray Kavukcuoglu, Demis Hassabis, Raia Hadsell, and Dharshan Kumaran. Vector-Based Navigation Using Grid-like Representations in Artificial Agents. *Nature*, 557(7705):429–433, May 2018. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-018-0102-6. URL http://www.nature.com/articles/s41586-018-0102-6.

Christopher J. Cueva and Xue-Xin Wei. Emergence of grid-like representations by training recurrent neural networks to perform spatial localization. In *International Conference on Learning Representations*, 2018. URL https://openreview.net/forum?id=B17JTOe0-.

Will Dorrell, Peter E. Latham, Timothy E. J. Behrens, and James C. R. Whittington. Actionable neural representations: Grid cells from minimal constraints. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=xfqDe72zh41.

Marianne Fyhn, Torkel Hafting, Alessandro Treves, May-Britt Moser, and Edvard I. Moser. Hippocampal Remapping and Grid Realignment in Entorhinal Cortex. *Nature*, 446(7132):190–194, March 2007. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature05601. URL http://www.nature.com/articles/nature05601.

Ruiqi Gao, Jianwen Xie, Xue-Xin Wei, Song-Chun Zhu, and Ying Nian Wu. On path integration of grid cells: group representation and isotropic scaling. *Advances in Neural Information Processing Systems*, 34:28623–28635, 2021.

Gily Ginosar, Johnatan Aljadeff, Liora Las, Dori Derdikman, and Nachum Ulanovsky. Are Grid Cells Used for Navigation? On Local Metrics, Subjective Spaces, and Black Holes. *Neuron*, 2023. ISSN 0896-6273. doi: 10.1016/j.neuron.2023.03.027. URL https://www.sciencedirect.com/science/article/pii/S0896627323002234.

Torkel Hafting, Marianne Fyhn, Sturla Molden, May-Britt Moser, and Edvard I. Moser. Microstructure of a spatial map in the entorhinal cortex. *Nature*, 436(7052):801–806, 2005. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature03721. URL http://www.nature.com/articles/nature03721.

Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pp. 448–456. pmlr, 2015.

Stefan Leutgeb, Jill K. Leutgeb, Alessandro Treves, May-Britt Moser, and Edvard I. Moser. Distinct Ensemble Codes in Hippocampal Areas CA3 and CA1. *Science*, 305(5688):1295–1298, August 2004. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1100265.

Daniel C. McNamee, Kimberly L. Stachenfeld, Matthew M. Botvinick, and Samuel J. Gershman. Flexible modulation of sequence generation in the entorhinal–hippocampal system. *Nature Neuroscience*, 24(6):851–862, Apr 2021. doi: 10.1038/s41593-021-00831-7.

J. O'Keefe and J. Dostrovsky. The Hippocampus as a Spatial Map. Preliminary Evidence from Unit Activity in the Freely-Moving Rat. *Brain Research*, 34(1):171–175, November 1971. ISSN 00068993. doi: 10.1016/0006-8993(71)90358-1. URL https://linkinghub.elsevier.com/retrieve/pii/0006899371903581.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.

Markus Pettersen, Frederik Rogge, and Mikkel Elle Lepperød. Learning place cell representations and context-dependent remapping. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. URL https://openreview.net/forum?id=7ESHFpqjNO.

Rylan Schaeffer, Mikail Khona, Tzuhsuan Ma, Cristobal Eyzaguirre, Sanmi Koyejo, and Ila Fiete. Self-supervised learning of representations for space generates multi-modular grid cells. *Advances in Neural Information Processing Systems*, 36:23140–23157, 2023.

Anirvan Sengupta, Cengiz Pehlevan, Mariano Tepper, Alexander Genkin, and Dmitri Chklovskii. Manifold-tiling localized receptive fields are optimal in similarity-preserving neural networks. *Advances in neural information processing systems*, 31, 2018.

Ben Sorscher, Gabriel C. Mel, Samuel A. Ocko, Lisa M. Giocomo, and Surya Ganguli. A unified theory for the computational and mechanistic origins of grid cells. *Neuron*, pp. S0896627322009072, October 2022. ISSN 08966273. doi: 10.1016/j.neuron.2022.10.003. URL https://linkinghub.elsevier.com/retrieve/pii/S0896627322009072.

Hanne Stensola, Tor Stensola, Trygve Solstad, Kristian Frøland, May-Britt Moser, and Edvard I. Moser. The Entorhinal Grid Map Is Discretized. *Nature*, 492(7427):72–78, December 2012. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature11649. URL http://www.nature.com/articles/nature11649.

Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020. doi: 10.1038/s41592-019-0686-2.

James C.R. Whittington, Timothy H. Muller, Shirley Mark, Guifen Chen, Caswell Barry, Neil Burgess, and Timothy E.J. Behrens. The Tolman-Eichenbaum Machine: Unifying Space and Relational Memory through Generalization in the Hippocampal Formation. *Cell*, 183(5):1249–1263.e23, November 2020. ISSN 00928674. doi: 10.1016/j.cell.2020.10.024. URL https://linkinghub.elsevier.com/retrieve/pii/S009286742031388X.

Dehong Xu, Ruiqi Gao, Wenhao Zhang, Xue-Xin Wei, and Ying Nian Wu. Conformal isometry of lie group representation in recurrent network of grid cells. In *NeurIPS 2022 Workshop on Symmetry and Geometry in Neural Representations*, 2022. URL https://openreview.net/forum?id=FszPdSkvGjz.

APPENDIX

A METHODS

All simulations were carried out using the matrix exponential in PyTorch (Paszke et al., 2019). To generate random, orthogonal matrices, we used the $ortho\_group$ functionality from the SciPy library (Virtanen et al., 2020), which uniformly samples matrices from the orthogonal group $O(N)$. For root-of-unity solutions, the arena size was 20×20 to reveal the full pattern of the representation, while for the similarity-preserving case the domain was $s \in [0, 5]$ in the non-spatial, and $x, y \in [-2, 2]$ for the spatial case, in line with the scale used for the desired Gaussian similarity function.

Large language models were used in writing this paper, with usage limited to improving writing and readability.

B COMMUTING GENERATORS PRODUCE PATH-INDEPENDENT REPRESENTATIONS

Because the entire spatial representation is furnished by the exponential map in equation 1, we can easily impose constraints on the representation by constraining the generators G_x and G_y. For example, Schaeffer et al. (2023)proposed that representations should be path-independent. In other words, the representation at a point, should not be contingent on the path travelled to get there.

In the exponential map formalism, this can be achieved exactly by demanding that the involved generators commute. Consider, for example, the representations at distinct points A, B, C, and D. Then, the path $A \to B \to D$ should give the same representation as the path $A \to C \to D$. The corresponding generated representations are

$$\mathbf{p}_{ABD} = e^{\Delta x_{BD} G_x + \Delta y_{BD} G_y} e^{\Delta x_{AB} G_x + \Delta y_{AB} G_y} \mathbf{p}_A,$$

and

$$\mathbf{p}_{ACD} = e^{\Delta x_{CD} G_x + \Delta y_{CD} G_y} e^{\Delta x_{AC} G_x + \Delta y_{AC} G_y} \mathbf{p}_A,$$

where $\mathbf{p}_A$ denotes the representation at A, while $\mathbf{p}_{ABD}$ denotes the representation at D, arrived at via B and so on. Note that to arrive at the final representations, we simply compose transforms from A to B/C, with transformations from B/C to D.

If G_x and G_y commute, so does any linear combination thereof. By the Baker-Campbell-Hausdorff formula, composite transformations may then be combined into a single transformation, i.e.

$$\mathbf{p}_{ABD} = e^{(\Delta x_{AB} + \Delta x_{BD}) G_x + (\Delta y_{AB} + \Delta y_{BD}) G_y} \mathbf{p}_A = \mathbf{p}_{ACD},$$

as the final state only depends on the displacement from the initial location, which is equal for both paths. Notably, if commutation is not satisfied, the generated state depends on the commutation relation of the generators, scaled by displacements along path segments.

In the non-commutative case, the path-dependence of the final representation arises due to the non-zero commutator of the generators G_x and G_y. Consider again the two paths, $A \to B \to D$ and $A \to C \to D$, with their respective representations:

$$\mathbf{p}_{ABD} = e^{\Delta x_{BD} G_x + \Delta y_{BD} G_y} e^{\Delta x_{AB} G_x + \Delta y_{AB} G_y} \mathbf{p}_A,$$

$$\mathbf{p}_{ACD} = e^{\Delta x_{CD} G_x + \Delta y_{CD} G_y} e^{\Delta x_{AC} G_x + \Delta y_{AC} G_y} \mathbf{p}_A.$$

When G_x and G_y do not commute, the Baker-Campbell-Hausdorff formula governs the combination of the exponential terms. Specifically, for matrices U and V,

$$e^U e^V = e^{U + V + \frac{1}{2}[U,V] + \frac{1}{12}([U,[U,V]] - [V,[U,V]]) + \cdots}.$$

Applying this to each path, the combined transformations for $\mathbf{p}_{ABD}$ and $\mathbf{p}_{ACD}$ differ due to the commutator terms introduced by the BCH expansion.

Let $U = \Delta x_{AB} G_x + \Delta y_{AB} G_y$ and $V = \Delta x_{BD} G_x + \Delta y_{BD} G_y$. Then

$$\mathbf{p}_{ABD} = e^U e^V \mathbf{p}_A = e^{U+V+\frac{1}{2}[U,V]+\cdots} \mathbf{p}_A.$$

Expanding $[U, V]$, we obtain

$$[U, V] = [\Delta x_{AB} G_x + \Delta y_{AB} G_y, \Delta x_{BD} G_x + \Delta y_{BD} G_y],$$

and when G_x and G_y do not commute, we have that

$$[U, V] = (\Delta x_{AB} \Delta y_{BD} - \Delta y_{AB} \Delta x_{BD})[G_x, G_y],$$

as $[G_x, G_y] = -[G_y, G_x]$. Notably, non-commuting generators G_x and G_y give rise to contributions that depend on the product of path segment displacements, and higher-order commutators will also contribute correction terms to the exponent. Similarly, for $\mathbf{p}_{ACD}$, the combined transformation is:

$$\mathbf{p}_{ACD} = e^W e^Z \mathbf{p}_A = e^{W+Z+\frac{1}{2}[W,Z]+\cdots} \mathbf{p}_A,$$

where $W = \Delta x_{AC} G_x + \Delta y_{AC} G_y$ and $Z = \Delta x_{CD} G_x + \Delta y_{CD} G_y$. The commutator $[W, Z]$ introduces terms analogous to $[U, V]$, but these terms now depend on products of path segment displacements, specific to the path $A \to C \to D$ (and higher-order terms). Thus, the representations at D, in general, depend on the specific path taken to get to it. In the commutative case, all commutators vanish, and the final representation depends only on the net displacement, which is path independent.

C From Generators to Representations

Given the form of the spatial representation in equation 1, we can rewrite a particular entry (i.e., a cell's spatial response) in a more insightful form. In particular, if we assume that $\mathbf{p}_0$ is a unit vector (for simplicity), and that generators are skew symmetric, commute, and are written in block diagonal form (as in equation 5), then

$$\boldsymbol{\alpha}(x, y) = e^{x\Sigma_x + y\Sigma_y} \boldsymbol{\alpha}_0,$$

with $\boldsymbol{\alpha} = R\mathbf{p}$, and $\boldsymbol{\alpha}_0 = R\mathbf{p}_0$ as before. Then, the matrix exponential itself now reduces to a block diagonal matrix, with 2×2 rotation matrices along the diagonal. This particular case has been studied previously by (Dorrell et al., 2023), and the resulting representation may stated as rotations in distinct 2D planes, where the action of a given block is

$$\boldsymbol{\alpha}^i = \begin{pmatrix} \cos(\Omega_i) & -\sin(\Omega_i) \\ \sin(\Omega_i) & \cos(\Omega_i) \end{pmatrix} \boldsymbol{\alpha}_0^i$$

where the uppercase indexes a block, meaning $i = 1, 2, ..., N/2$ (as there are $N/2$ blocks). Thus, $\boldsymbol{\alpha}^i \in \mathbb{R}^2$ is just a slice of the original transformed representation. Furthermore, $\Omega_i = x\lambda_{i,x} + y\lambda_{i,y}$ is a rotation angle that couples spatial location (or, for path integrating models, displacement) to eigenvalues of the generator matrices.

For a particular block, we have

$$\begin{aligned} \alpha_-^i &= \cos(\Omega_i)\alpha_{0,-}^i - \sin(\Omega_i)\alpha_{0,+}^i \\ \alpha_+^i &= \cos(\Omega_i)\alpha_{0,-}^i + \sin(\Omega_i)\alpha_{0,+}^i, \end{aligned}$$

where $\alpha_\pm^i$ is the entry corresponding to the eigenvalue $(+)$ of the ith block, and its conjugate $(-)$, respectively. As each entry is just a sum of two sinusoids, it can be written as

$$\begin{aligned} \alpha_-^i &= A_i \cos(x\lambda_{i,x} + y\lambda_{i,y} + \omega_i) \\ \alpha_+^i &= A_i \sin(x\lambda_{i,x} + y\lambda_{i,y} + \omega_i), \end{aligned}$$

where $A_i^2 = \alpha_{0,+}^2 + \alpha_{0,-}^2$ and $\omega_i = \arctan(-\alpha_{0,+}/\alpha_{0,-})$. Thus, before transformation by R, each entry is 2D plane wave, whose orientation and frequency is fixed by $\lambda_{i,x}$ and $\lambda_{i,y}$, and phase shifted by ω_i along the wave direction. Furthermore, the representation $\mathbf{p} = R^T\boldsymbol{\alpha}$ therefore consists of a mixture of plane waves.

D METRIC PRESERVATION

Consider the representation along a parametrized trajectory $\mathbf{r}(t) = (x(t), y(t))$, i.e.

$$\mathbf{p}(x(t), y(t)) = e^{x(t)G_x + y(t)G_y}\mathbf{p}_0.$$

In the representation, a line element can be written as $ds = |d\mathbf{p}|$, where

$$d\mathbf{p} = \left(\frac{\partial \mathbf{p}}{\partial x}\frac{dx}{dt} + \frac{\partial \mathbf{p}}{\partial y}\frac{dy}{dt}\right) dt$$

by the chain rule, meaning the length of a trajectory becomes

$$L = \int_0^S \sqrt{|d\mathbf{p}|^2} = \int_0^T \sqrt{\sum_{ij} g_{ij}\frac{dr_i}{dt}\frac{dr_j}{dt}}dt.$$

Comparing with the squared line element, we can then simply read off the induced metric g induced metric, as

$$g = -\begin{pmatrix} \mathbf{p}_0^T G_x^2 \mathbf{p}_0 & \mathbf{p}_0^T G_x G_y \mathbf{p}_0 \\ \mathbf{p}_0^T G_x G_y \mathbf{p}_0 & \mathbf{p}_0^T G_y^2 \mathbf{p}_0 \end{pmatrix},$$

since

$$\frac{\partial \mathbf{p}}{\partial r_i} = G_{r_i}\mathbf{p},$$

meaning

$$\frac{\partial \mathbf{p}}{\partial r_i}^T \frac{\partial \mathbf{p}}{\partial r_j} = \mathbf{p}^T G_{r_i}^T G_{r_j}\mathbf{p} = -\mathbf{p}_0^T G_{r_i} G_{r_j}\mathbf{p}_0$$

with $\mathbf{r} = (x, y)$, as the generator matrices are skew symmetric, and commute. We can simplify further by noting that

$$\begin{aligned} G_x^2 &= R^T\Sigma_x R R^T \Sigma_x R = R^T \Sigma_x^2 R \\ &= R^T D_x R, \end{aligned}$$

where D_x is a diagonal matrix, whose entries are the square of the generators' eigenvalue, $-\lambda_x^2$, as Σ_x is block diagonal with 2D skew-symmetric blocks (and eigenvalues are purely imaginary). Note that the same pattern holds for the other metric entries, with the off-diagonal product $G_x G_y = R^T D_{xy} R$ resulting in a diagonal matrix with products of eigenvalues on the diagonal, $-\lambda_x\lambda_y$. We may then write

$$g = -\begin{pmatrix} \boldsymbol{\alpha}_0^T D_x \boldsymbol{\alpha}_0 & \boldsymbol{\alpha}_0^T D_{xy} \boldsymbol{\alpha}_0 \\ \boldsymbol{\alpha}_0^T D_{xy} \boldsymbol{\alpha}_0 & \boldsymbol{\alpha}_0^T D_y \boldsymbol{\alpha}_0 \end{pmatrix}$$

with $\boldsymbol{\alpha}_0 = R\mathbf{p}_0$ as before. With this simplified form, the length of an induced path becomes

$$L = \int_0^T \sqrt{\sum_{i=1}^N \alpha_{0,i}^2(\lambda_{i,x}^2\dot{x}^2 + 2\lambda_{i,x}\lambda_{i,y}\dot{x}\dot{y} + \lambda_{i,y}^2\dot{y}^2)}\, dt$$

If we want our representation to preserve the flat metric, i.e., $g = \sigma^2 I$ for some σ, we need that

$$\sum_{i=1}^N \alpha_{0,i}^2\lambda_{ix}^2 = \sum_{i=1}^N \alpha_{0,i}^2\lambda_{iy}^2$$

and

$$\sum_{i=1}^N \alpha_{0,i}^2\lambda_{ix}\lambda_{iy} = 0.$$

Surprisingly, a rather straightforward solution exists: If we first introduce polar coordinates, $\lambda_{ix} = k_i\cos\phi_i,\ \lambda_{iy} = k_i\sin\phi_i$, we obtain the conditions

$$\begin{aligned} \sum_{i=1}^N \rho_i^2\cos^2(\phi_i) &= \sum_{i=1}^N \rho_i^2\sin^2(\phi_i) \\ \sum_{i=1}^N \rho_i^2\cos\phi_i\sin\phi_i &= 0, \end{aligned}$$

where $\rho_i = \alpha_i k_i$, or, equivalently

$$\sum_{i=1}^{N} \rho_i^2 \cos(2\phi_i) = -\sum_{i=1}^{N} \rho_i^2 \cos(2\phi_i)$$
$$\sum_{i=1}^{N} \rho_i^2 \sin(2\phi_i) = 0$$

which is readily derived by power-reduction and half-angle identities. Thus, we actually require

$$\sum_{i=1}^{N} \rho_i^2 \cos(2\phi_i) = 0 \quad \text{and} \quad \sum_{i=1}^{N} \rho_i^2 \sin(2\phi_i) = 0.$$

However, this just means that we need

$$\sum_{i=1}^{N} \rho_i^2 e^{2i\phi_i} = 0$$

as both the imaginary and real parts should vanish. Note, however, that for each ϕ_j, there is a conjugate $\phi_j^* = \phi_j + \pi$, as the eigenvalues of the generator matrices come in conjugate pairs. Therefore, the full expression can be written

$$\sum_{i=1}^{N/2} (\rho_i^2 + (\rho_i^*)^2) e^{2i\phi_j + 2\pi i} = 0,$$

but the conjugate phase shift does not impact the sum as $e^{2\pi i} = 1$. However, this enforces a requirement on our choice of ρ_i, which we will take to be equal to ρ_i^* going forward. Note also that we already restricted ourselves to the case where N is even.

The simplest case is when ρ_i is a shared quantity, i.e. when $\rho_i = \rho$, as we only need

$$Z = \sum_{j=1}^{N/2} e^{2i\phi_j} = 0.$$

Notably, this requirement holds for any set of *roots of unity*, so if we simply choose

$$\phi_j = \pi \frac{j}{N}, \quad j = 0, 1, ..., N-1$$

then the representation preserves the Euclidean metric! However, we can find a broader class of solutions by noting that a rotation of a set of roots of unity, i.e., letting $\phi_i \to \varphi_i + \pi \frac{i}{N}$ for a shared phase φ_i, maintains $Z = 0$ as the sum of phasors still cancel. Furthermore, the radius of a given set of roots of unity does not matter, as long as the roots lie on the same ring in the complex plane. Finally, any linear combination of such sets also sums to zero, as each set of roots of unity sums to zero individually. Therefore, a solution may be of the form

$$Z = \sum_{j=1}^{J} \rho_j^2 e^{2i\varphi_j} \sum_{m=0}^{M_j - 1} e^{2\pi i \frac{m}{M_j}}.$$

In other words, for each radius ρ, there can be multiple rotated sets of roots of unity, each with its own rotational symmetry.

Comparing with the explicit form of the representation in Appendix C, this result is reminiscent of the modular organization of grid cells in the Entorhinal Cortex (Hafting et al., 2005; Stensola et al., 2012), which are organized in distinct modules with different grid spacings (a particular ρ), pattern orientations (a phase offset φ), and a pattern symmetry (a shared M_j).

E SIMILARITY FUNCTION DERIVATION

To derive an explicit form of the similarity between different representations, we start from the similarity function in equation 4, and demand that generators are skew symmetric, and commute by taking

$$G_x = R^T \Sigma_x R \quad \text{and} \quad G_y = R^T \Sigma_y R,$$

where R is orthogonal and shared by G_x and G_y. We then make use of another property of the matrix exponential, namely that for a similarity transformation $P^{-1}AP$, for some matrices A and P, we may write

$$e^{P^{-1}AP} = P^{-1} e^A P.$$

Using the block diagonal form, the similarity in equation 4 may therefore be written as

$$\begin{aligned} C(\Delta x, \Delta y) &= \left(R \frac{\mathbf{p}_0}{|\mathbf{p}_0|} \right)^T e^{\Delta x \Sigma_x + \Delta y \Sigma_y} \left(R \frac{\mathbf{p}_0}{|\mathbf{p}_0|} \right) \\ &= \boldsymbol{\alpha}_0^T e^{\Delta x \Sigma_x + \Delta y \Sigma_y} \boldsymbol{\alpha}_0, \end{aligned}$$

where we have dubbed $\boldsymbol{\alpha}_0 \equiv R \frac{\mathbf{p}_0}{|\mathbf{p}_0|}$ for legibility, and $\sum_i \alpha_{0,i}^2 = 1$ as R is orthogonal. Notice that the exponent matrix is still skew symmetric, with the same $2D$ block structure as before. Expanding the matrix exponential, one finds that even powers of this matrix are diagonal, while odd powers are skew symmetric. As the quadratic form vanishes under a skew symmetric matrix, we are left with a sum over even powers of the form

$$C(\Delta x, \Delta y) = \boldsymbol{\alpha}_0^T \left(\sum_n^\infty \frac{(-1)^n}{(2n)!} D^{2n} \right) \boldsymbol{\alpha}_0, \tag{9}$$

where D is a diagonal matrix with entries $d_{ii} = \lambda_{i,x} \Delta x + \lambda_{i,y} \Delta y$, with $\lambda_{i,x}$ being the imaginary part of the i-th eigenvalue of G_x, and so on. However, the matrix sum in equation 9 is nothing but a diagonal matrix with cosine entries along the diagonal (by the Taylor expansion of the cosine). Therefore, the similarity admits a particularly simple form

$$C(\Delta x, \Delta y) = \sum_i^N \alpha_{0,i}^2 \cos(\lambda_{i,x} \Delta x + \lambda_{i,y} \Delta y).$$

F DESIGNING SPATIAL SIMILARITY FUNCTIONS

We found in general that the similarity function equation 6 may be written as a weighted sum of cosines. However, it is yet unclear what the representational similarity could, or should, be. To untangle this question, we can first rewrite it in polar coordinates,

$$C = \sum_i \alpha_{0,i}^2 \cos(k_i r \cos(\theta - \phi_i)),$$

where we have introduced $x = r \cos\theta$, $y = r \sin\theta$ and $\lambda_{i,x} = k_i \cos\phi_i$, $\lambda_{i,y} = k_i \sin\phi_i$. Using the Jacobi-Anger expansion, we may further write

$$\begin{aligned} \cos(z\cos(\omega)) &= J_0(z) + 2\sum_{n=1}^\infty (-1)^n J_{2n}(z) \cos(2n\omega) \\ &= J_0(z) + 2\sum_{n=1}^\infty (-1)^n \Re\left\{ J_{2n}(z) e^{2in\omega} \right\}, \end{aligned}$$

where $J_n(z)$ is the nth Bessel function of the first kind. If the sum is well-behaved, we can use this identity to rewrite the similarity function as

$$C(r,\theta) = \sum_j \alpha_{0,j}^2 J_0(k_j r) + 2\sum_{n=1}^\infty (-1)^n \Re\left\{ e^{2in\theta} \sum_j \alpha_{0,j}^2 J_{2n}(k_j r) e^{2in\phi_j} \right\},$$

i.e., as a purely radial contribution (the sum over J_0) plus a mixed radial–head-direction–dependent part. At this point, no further simplification is possible unless we impose additional constraints on the representation.

However, the inner sum

$$\sum_j \alpha_{0,j}^2 J_{2n}(k_j r) e^{2in\phi_j}$$

closely resembles the phasor sum obtained in Appendix D. In particular, if the representation is *metric-preserving* (so that eigenvalues are distributed in discrete root-of-unity constellations) and we assume $\alpha_{0,j}$ is constant on each ring, we can set

$$\sum_j \alpha_{0,j}^2 J_{2n}(k_j r) e^{2in\phi_j} = \sum_j \alpha_{0,j}^2 J_{2n}(k_j r)\, e^{2in\varphi_j} \sum_{m=0}^{M_j-1} e^{2\pi imn/M_j},$$

where the inner sum is a geometric series of the form

$$\sum_{m=0}^{M_j-1} e^{2\pi imn/M_j} \;=\; \begin{cases} M_j, & n = \ell M_j \quad (\ell \in \mathbb{Z}), \\ 0, & \text{otherwise.} \end{cases}$$

Thus, if the representation contains multiple symmetries M_j, the lowest-order head-direction–dependent term appears at $n = M_{\min}$. If all rings share the same symmetry order M, then only terms with $n = \ell M$ survive. For large M, i.e., a near-uniform distribution of eigenvalues around the circle, the similarity becomes approximately head-direction independent.

We can further suppress low-order angular terms by also requiring *orientation offsets* to form a root-of-unity constellation. That is, if

$$\sum_j \alpha_{0,j}^2 J_{2n}(k_j r) e^{2in\phi_j} = \sum_j \alpha_{0,j}^2 J_{2n}(k_j r) \sum_l e^{2in\varphi_l} \sum_{m=0}^{M-1} e^{2\pi imn/M},$$

and the orientations themselves satisfy

$$\sum_l e^{2in\varphi_l} = \sum_{l=0}^{N-1} e^{2\pi iln/N},$$

then only terms with $n = zN$ and simultaneously $n = pM$ survive (for $z, p \in \mathbb{Z}$). The most head-direction–independent representation arises when M and N are coprime, in which case angular terms appear only at multiples of MN.

In this case, the similarity is just

$$C(r,\theta) = \sum_j \alpha_{0,j}^2 J_0(k_j r) + 2MN \sum_{\ell=1}^{\infty} (-1)^{\ell NM} \alpha_{0,j}^2 J_{2\ell MN}(k_j r) \cos(2\ell MN\theta).$$

However, when $0 < z << \sqrt{\gamma+1}$ then

$$J_\gamma(z) \approx \frac{1}{\Gamma(\gamma+1)} \left(\frac{z}{2}\right)^\gamma$$

meaning that for large NM, the correlation is approximately head direction independent for a large range of displacements, r, and takes the following form

$$\begin{aligned} C(r,\theta) &\approx \hat{C}(r) \\ &= \sum_j \alpha_{0,j}^2 J_0(k_j r). \end{aligned}$$

Thus, when the eigenvalues that generate the representation are modularly arranged in constellations of roots of unity, the resulting similarity function is approximately radial. Also, for a single set of roots of unity, the similarity function is approximately $J_0(kr)$. However, if more rings are included, the final approximate expression is a Fourier-Bessel series in $k_j r$! Thus, the metric-preserving

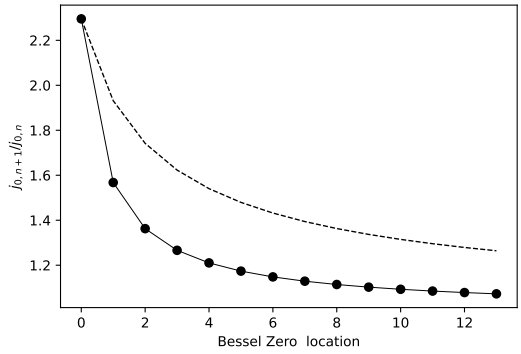

Figure 4: Ratio of subsequent zeros of the Bessel function J_0 (large dots), alongside a cumulative average (dashed line).

representation can approximate a range of functions for small/intermediate r, if k_j are proportional to the zeros of J_0 (which is how a Fourier-Bessel series is constructed).

Intriguingly, the ratio between subsequent, low-order zeroes of the Bessel function falls in the same variability range as that observed experimentally for grid cell modules (Stensola et al., 2012). The average ratio is often reported as being close to $\sqrt{2}$. For Bessel zeroes, however, the cumulative average depends on the number of zeroes included, but for a small number of zeroes the mean is close to this particular value, which is shown in Fig. 4. Thus, the ratios of grid cell spacings could conceivably be related to zeroes of the Bessel function.

G SIMILARITY-PRESERVING SPATIAL REPRESENTATIONS

If we drop the requirement that the spatial representation should preserve the metric of space, we can instead consider similarity-preserving representations. As a concrete example, we consider the case where generator eigenvalues are sampled from a normal distribution, as described in 2.5. Then, the representation is approximately similarity-preserving, and the spatial similarity is approximately Gaussian, as shown in the non-spatial case. To demonstrate that this generalizes to spatial representations, we ran a simulation where eigenvalues of generators G_x and G_y with $N = 256$ units were sampled according to a normal distribution $\mathcal{N}(0, 2)$. The result is shown in Fig. 5, where example ratemaps of a similarity-preserving spatial representation is shown. Notably, units do not display a strong periodic tuning, as with the low-order roots-of-unity solutions for metric preservation. However, representations are in some cases tuned to particular locations, reminiscent of Hippocampal place fields. When viewed in light of the fact that place cells are known to respond to spatial, as well as nonspatial cues, such as room smell (Anderson & Jeffery, 2003), it could be interesting to model conjunctive representations of space and context, similar to (Pettersen et al., 2024), using the context-dependent model in equation 7.

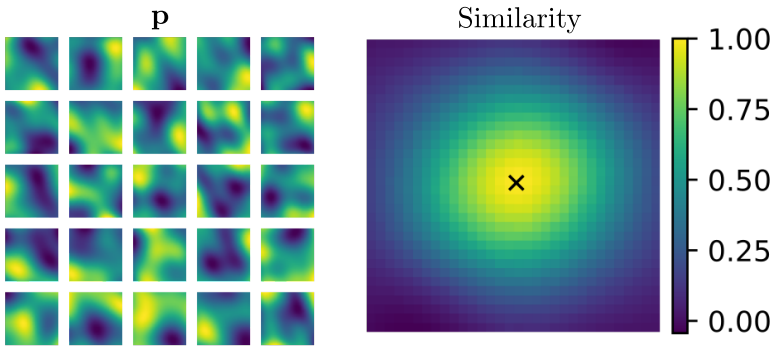

Figure 5: Similarity-preserving spatial representations. The left-hand side shows example ratemaps of a model whose eigenvalues are sampled from a normal distribution, such that the similarity function is approximately Gaussian. The resulting similarity, relative to the origin, is shown on the right-hand side.