Group Equivariant Sparse Coding

Christian Shewmake[1,2,3(✉)], Nina Miolane[3], and Bruno Olshausen[1,2]

[1] University of California Berkeley, Berkeley, CA 94720, USA
[2] The Redwood Center for Theoretical Neuroscience, Berkeley, CA 94720, USA
[3] University of California Santa Barbara, Santa Barbara, CA 93106, USA
shewmake@berkeley.edu

Abstract. We describe a sparse coding model of visual cortex that encodes image transformations in an equivariant and hierarchical manner. The model consists of a group-equivariant convolutional layer with internal recurrent connections that implement sparse coding through neural population attractor dynamics, consistent with the architecture of visual cortex. The layers can be stacked hierarchically by introducing recurrent connections between them. The hierarchical structure enables rich bottom-up and top-down information flows, hypothesized to underlie the visual system's ability for perceptual inference. The model's equivariant representations are demonstrated on time-varying visual scenes.

Keywords: Equivariance · Sparse coding · Generative models

1 Introduction

Brains have the remarkable ability to build internal models from sensory data for reasoning, learning, and prediction to guide actions in dynamic environments. Central to this is the problem of *representation*—i.e., how do neural systems construct internal representations of the world? In the Bayesian view, this requires a generative model mapping from a latent state space to observations, along with a mechanism for inferring latent states from sensory data. Thus, understanding the causal structure of the natural world is essential for forming internal representations. But what is the causal structure of the natural world? Natural images contain complex *transformation groups* that act both on objects and their parts. Variations in object pose, articulation of its parts, even lighting and color changes, can be described by the actions of groups. Additionally, these variations are hierarchical in nature: scenes are composed of objects, objects are composed of parts in relative poses, and so on down to low-level image features. A transformation at the level of an object propagates down the compositional hierarchy, transforming each of its component parts correspondingly. Finally, object parts and sub-parts can undergo their own independent transformations. These variations carry important information for understanding and meaningfully interacting with the world. Thus, a rich compositional hierarchy that is *compatible with group actions* is essential for forming visual representations (Fig. 1).

F. Nielsen and F. Barbaresco (Eds.): GSI 2023, LNCS 14071, pp. 91–101, 2023.
https://doi.org/10.1007/978-3-031-38271-0_10

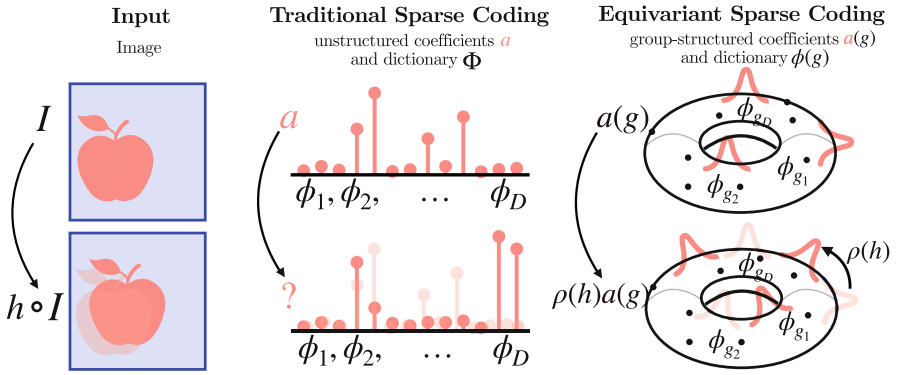

Fig. 1. Traditional vs Equivariant sparse coding as image I is transformed by action h.

Our contribution. We establish a novel Bayesian model for forming representations of visual scenes with equivariant hierarchical part-whole relations by proposing a group-equivariant extension of hierarchical *sparse coding* [7].

2 Background: Sparse Coding for Visual Representations

Sparse coding was originally proposed as a model for how neurons in primary visual cortex represent image data coming from the retina. In contrast to the feedforward cascade of linear filtering followed by point-wise nonlinearity commonly utilized in deep learning architectures, sparse coding uses *recurrent dynamics* to infer a sparse representation of images in terms of a learned dictionary of image features. When trained on natural images, the learned dictionary resembles the oriented, localized, and bandpass receptive fields of neurons in primary visual cortex (area V1) [7].

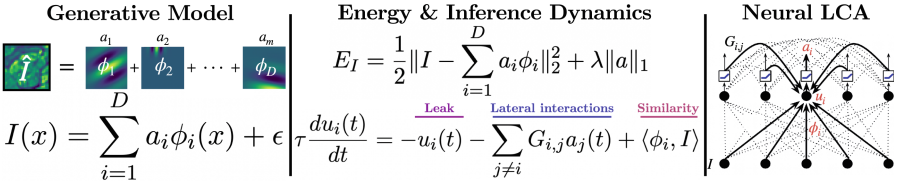

Fig. 2. (left) Generative model, (center) Energy function where $\|\cdot\|_2$ denotes the Euclidean norm, $\|\cdot\|_1$ denotes the ℓ_1 norm, and λ is a regularization parameter controlling the sparsity of a. u_i is the internal state of neuron i, $G_{i,j} = \langle \phi_i, \phi_j \rangle$ models neuronal interactions, and $a(t) = \sigma(u(t))$, where σ is a nonlinearity. (right) LCA circuit model

Generative Model. Sparse coding assumes that natural images are described by a linear generative model with an overcomplete dictionary and additive Gaussian noise $\epsilon(x)$ [7], shown in Fig. 2 (left). Here, the image I is represented as a function $I : X \rightarrow \mathbb{R}$, specifically as a vector in the space $L_2(X)$ of square-integrable functions with compact support $X \subset \mathbb{R}^2$. Computationally, the support is discretized as an image patch with n pixels, so that $I \in \mathbb{R}^n$. The dictionary Φ comprises D elements: $\Phi = \{\phi_1, \ldots, \phi_D\}$, with each $\phi_i \in L_2(X)$, for $i \in \{1, ..., D\}$. The size of the dictionary D is typically chosen to be overcomplete, i.e. larger than the image patch dimension n. The coefficients $a = [a_1, \ldots, a_D] \in \mathbb{R}^D$ form the *representation of image* I.

Energy & Inference Dynamics. Given a dataset, sparse coding attempts to find a dictionary Φ and a latent representation $a \in \mathbb{R}^D$ for each image in the dataset such that, in expectation, neural activations are maximally sparse and independent. Sparsity is promoted through the use of an i.i.d.prior over a with scale parameter λ, with the form of the prior chosen to be peaked at zero with heavy tails compared to a Gaussian (typically Laplacian). Finding the optimal representation a is accomplished by maximizing the posterior distribution $P(a|I, \Phi)$ via minimization of the energy function E_I in Fig. 2 (center).
One particularly effective method for minimizing E_I with a clear cortical circuit implementation is the *Locally Competitive Algorithm* (LCA) [9]. In LCA, inference is carried out via the temporal dynamics of a population of D neurons. Each neuron is associated with a dictionary element i, and its internal state is represented by a coefficient $u_i(t)$. The evolution of the neural population state is governed by the dynamics specified in Fig. 2 (center). The gram matrix, $G_{i,j} = \langle \phi_i, \phi_j \rangle$, specifies the interaction between neuron i and j. In neurobiological terms, this corresponds to the excitatory and inhibitory interactions mediated by horizontal connections among V1 neurons. The notation $\langle ., . \rangle$ refers to the inner-product between functions in $L_2(X)$, $\langle \phi_i, \phi_j \rangle = \int_X \phi_i(x)\phi_j(x)dx$. The activations, interpreted as instantaneous neural firing rates, are given by a nonlinearity applied to the internal state: $a_j(t) = \sigma(u_j(t))$, with $\sigma(u) = \frac{u - \alpha\lambda}{1+e^{-\gamma(u-\lambda)}}$, similar to a smoothed ReLU function with threshold λ and hyperparameters α and γ. At equilibrium, the latent representation of image I is given by $\hat{a} = \arg\min_a E_I$.

Dictionary Learning. The dictionary Φ is adapted to the statistics of the data by minimizing the same energy function E_I averaged over the dataset. This is accomplished by alternating gradient descent on E. Given a current dictionary Φ, along with a batch of images and their inferred latent representations, $\hat{a}$, the dictionary is updated with one gradient step of E with respect to Φ, averaged over the batch.

3 Group Equivariant Sparse Coding

Missing in the current formulation of sparse coding is the mathematical structure to support reasoning about hierarchical object transformations in visual

scenes. This limits its utility in both unsupervised learning and mechanistic models of visual cortex. Here we address this problem by explicitly incorporating group equivariant and hierarchical structure into the sparse coding model. Prior work has explored imposing topological relations between dictionary elements by establishing implicit neighborhood relations during training through co-activation penalties [6], or explicitly coupling steerable pairs or n-tuples of dictionary elements [8]. More recent work in Geometric Deep Learning (GDL) has introduced several group equivariant architectures, for example through the use of group convolutions [3,4]. However, these models are feedforward, lacking mechanisms for hierarchical inference or rich top-down and bottom up flows. Aside from [1], these models lack mechanisms for hierarchical part-whole relations.

We explore the implications of inheriting the dictionary's geometric structure through *group actions*. In particular, we propose a model in which each dictionary element is generated by an action of g on a canonical dictionary element, as shown in Fig. 3 (right). For example, the group G of 2D rotations acts on the 2D domain X, inducing a natural action on the space of images in $L_2(X)$ defined over X. We refer the interested reader to [5] for mathematical details on groups and group actions.

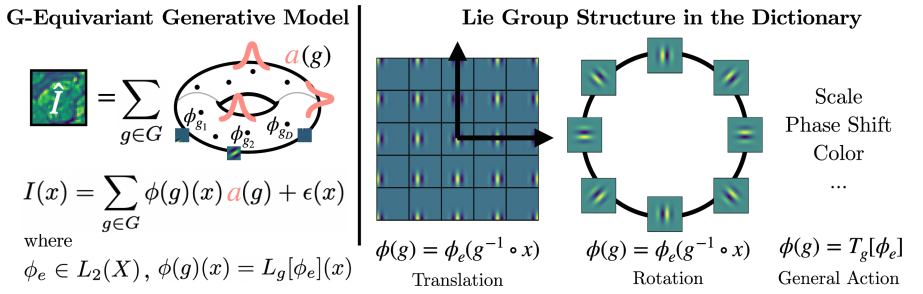

Fig. 3. (left) Geometric generative model, (right) Lie group actions relate dictionary elements. Here, e is the identity element of G, and the canonical dictionary element is $\phi_e \in L_2(X)$. Additionally, L is a linear group action of G in the space of functions on the domain $L_h[f](x) = f(h^{-1}x)$.

3.1 Geometric Generative Model

This perspective enables us to rewrite the sparse coding generative model as:

$$I(x) = \sum_{g \in G} \phi(g)(x)\, a(g) + \epsilon(x), \tag{1}$$

where both the dictionary elements $\phi(g)$ and the scalar coefficients $a(g)$ are indexed with group elements, i.e. "coordinates" in G. In other words, images are (1) generated by linear combinations of dictionary elements ϕ, where (2)

each dictionary element has an explicit coordinate g in the group. The latent representation a is now a scalar field over the group, $a : G \to \mathbb{R}$, illustrated in Figs. 3 (left) and 1 (right). Intuitively, this perspective gives an explicit geometric interpretation of both the dictionary Φ and latent representation a in sparse coding, and thus a route toward modeling transformations which was implicit in the unstructured vector representation.

3.2 Geometric Inference and LCA

The geometric perspective of sparse coding above allows us to rewrite the LCA dynamics. Specifically, each neuron is now associated with a group element g, with internal state $u(g)$. The LCA dynamics are typically computationally expensive to compute due to the prohibitive size of the neural interaction matrix $G_{g,h} = \langle \phi(g), \phi(h) \rangle$. However this term can now be written as a group convolution with a ϕ_e-dependent kernel w, leading to a **symmetric, local wiring rule** between neurons and efficient computation during inference that is readily parallelized on GPUs. Hence, we propose a new, provably equivariant inference method—*Geometric LCA*, where $*$ denotes group convolution:

$$\dot{u}(g)(t) = -u(g)(t) - [w * a(t)](g) + \langle \phi(g), I \rangle \tag{2}$$

Box 1. Isometry and the Derivation of Geometric LCA

Lemma 1 Consider a function $f \in L_2(X)$ and a dictionary element $\phi(g) \in L_2(X)$ indexed by $g \in G$. If the action of $h \in G$ is isometric on the domain X, then, $\forall h \in G$, we have:

$$\langle L_h[\phi(g)], f \rangle = \langle \phi(g), L_{h^{-1}}[f] \rangle$$

Proof. We have: $\langle L_h[\phi(g)], f \rangle$

$$= \int_X L_h[\phi(g)](x) f(x) dx$$
$$= \int_X L_{hg}[\phi_e](x) f(x) dx \quad \text{by def. of } \phi(g)$$
$$= \int_X \phi_e((hg)^{-1}x) f(x) dx \quad \text{by def. of } L$$
$$= \int_X \phi_e(g^{-1}h^{-1}x) f(x) dx$$

$x \leftarrow hx$, h action isometric: $d(hx) = dx$

$$= \int_X \phi_e(g^{-1}x) f(hx) dx$$
$$= \int_X \phi(g)(x) L_{h^{-1}}[f](x) dx$$
$$= \langle \phi(g), L_{h^{-1}}[f] \rangle.$$

This last step leads to the following LCA dynamics.

Proposition 1: Geometric LCA The LCA dynamics have the following geometric formulation

$$\tau \dot{u}(g)(t) = -u(g)(t) - [w * a(t)](g) + \langle \phi(g), I \rangle$$

where $*$ denotes a group convolution.

Proof. Consider the interaction term in the LCA dynamics: $\sum_{h \in G,\ h \neq g} G_{g,h} a(h)(t)$

$$= \sum_{h \in G,\ h \neq g} \langle \phi(g), \phi(h) \rangle a(h)(t) \quad \text{by def. of } G$$
$$= \sum_{h \in G,\ h \neq g} \langle L_h^{-1}[\phi(g)], \phi_e \rangle a(h)(t) \quad \text{Lemma 1}$$
$$= \sum_{h \in G,\ h \neq g} \langle \phi(h^{-1}g), \phi_e \rangle a(h)(t)$$
$$= \sum_{h \in G,\ h \neq g} w(g^{-1}h) a(h)(t)$$
$$= [w * a(t)]\,(g),$$

where we define $w(g) \coloneqq \langle \phi(g^{-1}), \phi_e \rangle$ for $g \neq e$ and $w(e) = 0$.

Equivariance of Inference and LCA. Next, we demonstrate that the solutions $I \to a$ obtained from the LCA dynamics are equivariant. First, we say a map $\psi : X \to Z$ is *equivariant* to a group G if $\psi(L_g x) = L'_g \psi(x)\ \forall g \in G$, with L_g, L'_g representations of G on X and Z respectively. For clarity of exposition, L is defined as a group action of G on the space $L_2(X)$ via domain transformations $L_g[f](x) = f(g^{-1}x)$, and the action L' of G is defined on the space $L_2(G)$ of square integrable functions from G to $\mathbb{R}$, defined as:

$$L'_h(a)(g) = a(h^{-1}g), \quad \forall g \in G,\ \forall h \in G,\ \forall a \in L_2(G),$$

where $h^{-1}g$ refers to the group composition of two group elements. First, we show that solutions of the ordinary differential equation (ODE) defining the LCA dynamics exist and are unique. Consider the initial value problem below, where f denotes the LCA dynamics:

$$\text{ODE}(I) : \begin{cases} \dot{u}(g,t) = f(u(g,t), I) & \forall g \in G, t \in \mathbb{R}_+, \\ u(g,0) = 0 & \forall g \in G. \end{cases} \tag{3}$$

Proposition 1 (Existence and Uniqueness of LCA Solutions). *Given an image I, the solution of* ODE(I) *exists and is unique. We denote it with u^I.*

Proof. The Cauchy-Lipschitz theorem (Picard-Lindelöf theorem) states that the initial value problem defined by $ODE(I)$ has a unique solution if the function f is (i) continuous in t and (ii) Lipschitz continuous in u, where:

$$f(u(g,t), I) = \frac{1}{\tau}\left(-u(g)(t) - [w * a(t)](g) + \langle \phi(g), I \rangle\right) \tag{4}$$

The continuity in t stems from the fact that a and u are continuous. We prove that $f(u, I)$ is Lipschitz continuous in u, i.e. that $\frac{\partial f}{\partial u}(u, I)$ is bounded. Observe that the derivatives of the first and third terms are bounded. The second term is a convolution composed with a smooth, ReLU-like nonlinearity. As convolutions are bounded linear operators, the question reduces to whether derivative of the nonlinearity $\frac{\partial \sigma}{\partial u}$ is bounded, which indeed holds. Thus solutions exist and are unique. Using this fact, we show that the solution of the dynamics transforms equivariantly with image transformations. Let u^I be the unique solution of ODE(I). Similarly, let $u^{L_h[I]}$ be the unique solution of:

$$\text{ODE}(L_h[I]) : \begin{cases} \dot{u}(g,t) = f(u(g,t), L_h[I]) & \forall g \in G, t \in \mathbb{R}_+, \\ u(g,0) = 0 & \forall g \in G. \end{cases}$$

Proposition 2: Equivariance of LCA Inference Dynamics Take $h \in G$. The solutions of the LCA dynamics $ODE(I)$ and $ODE(L_h[I])$ are related by $u^{L_h I} = L'_h(u^I)$. Since $a(g) = \sigma(u(g))$, it follows that: $a^{L_h I} = L'_h(a^I)$.

Proof Take $h \in G$ and define $v(g,t) := u^I(h^{-1}g,t), \forall g, \forall t$. We show v verifies ODE($L_h[I]$). First, we verify that v satisfies the initial conditions: $v(g,0) = u^I(h^{-1}g,0) = 0, \ \forall g \in G$. Next, we verify that v satisfies ODE($L_h[I]$) $\ \forall g, \forall t$.

$$
\begin{aligned}
\tau \dot{v}(g,t) &= \frac{\partial}{\partial t}[\tau u^I(h^{-1}g,t)] \quad \text{(definition of } v\text{)} \\
&= -u^I(h^{-1}g,t) - \sum_{g' \in G} w((h^{-1}g)^{-1}g') \cdot \sigma(u(g')) + \langle \phi(h^{-1}g), I \rangle \\
&= -v(g,t) - \sum_{g' \in G} w(g^{-1}hg') \cdot \sigma(u(g')) + \langle \phi(g), L_h[I] \rangle \quad \text{(Lemma 1)} \\
&= -v(g,t) - \sum_{g' \in G} w(g^{-1}g') \cdot \sigma(u(h^{-1}g')) + \langle \phi(g), L_h[I] \rangle \quad (g' \leftarrow h^{-1}g') \\
&= -v(g,t) - \sum_{g' \in G} w(g^{-1}g') \cdot \sigma(v(g')) + \langle \phi(g), L_h[I] \rangle \quad \text{(definition of } v\text{)} \\
&= f(v(g,t), L_h[I]) \quad \text{(definition of ODE } (L_h[I])\text{)}.
\end{aligned}
$$

Thus, v is a solution of ODE($L_h[I]$), and, by uniqueness, $v(g,t) = u^{L_h[I]}(g,t) \ \ \forall g, \forall t$. Therefore, $u^I(h^{-1}g,t) = u^{L_h[I]}(g,t) \ \ \forall g, \forall t$, and $a^{L_h I} = L'_h(a^I)$ as well. Thus, the LCA inference dynamics are equivariant to global image transformations.

3.3 Equivariances of the Generative Model

Here, we show that the generative model, that is, the function $f : a \rightarrow I$ that maps coefficients to images, is also equivariant. There are three types of equivariance important for representing transformations in natural scenes: global, part/local, and hierarchical. Here we define these three types and prove that the generative model is indeed equivariant in these important ways.

Global Equivariance. Traditionally, work on group equivariant neural networks (e.g. GCNNs [3,4]) has focused on global equivariance, i.e. equivariance to a group action L on the domain of the input function. In Box 2, we show that the geometric form of the sparse coding model is globally equivariant. However, transformations of natural scenes typically involve actions on objects and parts at different levels of the hierarchy. That is, transformations of an object at a higher level of the hierarchy should propagate down compatibly with its parts. In the context of equivariant sparse coding, the generative model explicitly decomposes the scene into primitive parts—the first-level dictionary elements. That is, if an

image I is composed of M objects I_1, ..., I_M then

$$\begin{aligned} I(x) &= I_1(x) + ... + I_M(x) \\ &= \sum_{g \in G} \phi(g)(x) a_1(g) + ... + \sum_{g \in G} \phi(g)(x) a_M(g) \\ &= \sum_{g \in G} \phi(g)(x) \left(a_1(g) + ... + a_M(g)\right). \end{aligned}$$

In the context of this generative model, we can define two additional notions that are essential for natural scene decompositions—*local* and *hierarchical* equivariance. We prove that the generative model is indeed equivariant in these two additional important ways.

Local Equivariance. Using the decomposition above, we define *local actions* of G on the space of images $L_2(X)$ as:

$$L_h^{(1)}[I] = L_h[I_1] + ... + I_M, \quad \forall h \in G, I \in L_2(X)$$

$L^{(1)}$ only acts on image part 1, represented by image I_1, and likewise on image part m via $L^{(m)}$. We now prove that these local actions are indeed group actions.
Proof $L^{(1)}$ is a group action. The proof for $L^{(2)}$ follows.
(i) *Identity* : $L_e^{(1)}[I] = I$.
(ii) *Closure* : $L_{h'h}^{(1)}[I] = L_{h'h}[I_1] + ... + I_M = L_{h'}\left[L_h[I_1]\right] + ... + I_M = L_{h'}^{(1)}\left[L_h^{(1)}[I]\right]$.
Here, we have used the definition of $L^{(1)}$ and the fact that L is a group action. Similarly, we can define local actions $L'^{(m)}$ on the space $L_2(G)$ of coefficients a_m corresponding to image part I_m. By the linearity of the generative model f, a local action in the space of coefficients yields a local action in the image space, as shown in Box 2, Proposition 3.

Hierarchical Equivariance. The properties of global and local equivariance naturally give rise to the hierarchical equivariance of the new generative model. In other words, when a transformation is applied at the level of an object I (e.g. the whole scene), transformations propagate down compatibly to its parts (e.g. $I_1, ..., I_M$). This hierarchical transformation is directly reflected in actions on the latent coefficients for an object a and its parts $a_1, ..., a_M$. See Box 2, Proposition 5. Thus, a hierarchy of transformations in the scene is equivalent to a hierarchy of transformations in the internal neural representation.

Box 2. Generative Model: Global, Local, & Hierarchical Equivariance

Proposition 2: Global The generative model in Eq. 1, $I = f(a)$ is globally G-equivariant, i.e. for all $h \in G$, we have:

$$f(L'_h(a)) = L_h(f(a)) \tag{5}$$

Proof Take $h \in G$. We have:

$$\begin{aligned} L_h[f(a)] &= L_h\left[\sum_{g\in G} \phi(g)a(g)\right] \\ &= \sum_{g\in G} L_h\left[L_g[\phi_e]\right] a(g) \\ &= \sum_{g\in G} \phi(hg)a(g) \\ &= \sum_{'\in G} \phi(g)a(h^{-1}g) \quad (g \leftarrow h^{-1}g) \\ &= \sum_{g\in G} \phi(g)L'_h(a)(g). \end{aligned}$$

Thus the model is globally G-equivariant.

Proposition 3: Part/Local Consider the linear model f, where $a = a_1 + ... + a_M$. f is locally G-equivariant, i.e. $\forall h \in G$, for $m \in \{1, 2, ..., M\}$

$$f\left(L'^{(m)}_h(a)\right) = L^{(m)}_h[f(a)] \tag{6}$$

Proof We have

$$\begin{aligned} f\left(L'^{(1)}_h(a)\right) &= f\left(L'_h(a_1) + ... + a_M\right) \\ &= f\left(L'_h(a_1)\right) + ... + f(a_M) \\ &= f\left(L'_h(a_1)\right) + ... + I_M \\ &= L_h\left[I_1\right] + ... + I_M \quad \text{(by 5)} \\ &= L^{(1)}[I] \quad \text{(definition of } L^{(1)}\text{)}. \end{aligned}$$

Shown for $m = 1$, this property holds for all m, thus the model is locally G-equivariant.

Proposition 4: Hierarchical Consider the linear model f, where $a = a_1 + ... + a_M$. For all $h, h' \in G$, $m \in \{1, 2, ..., M\}$ we have:

$$f\left(L'_h\left(L'^{(m)}_{h'}(a)\right)\right) = L_h\left[L^{(m)}_{h'}\left[f(a)\right]\right]$$

Proof Directly from global and local cases:

$$\begin{aligned} f\left(L'_h\left(L'^{(m)}_{h'}(a)\right)\right) &= L_h\left[f\left(L'^{(m)}_{h'}(a)\right)\right] \\ &= L_h\left[L^{(m)}_{h'}\left[f(a)\right]\right]. \end{aligned}$$

Thus, f is *hierarchically* G-equivariant.

3.4 Constructing a Hierarchical Generative Model

Finally, the equivariant sparse coding layers can be composed hierarchically, where first-level activations are describable in terms of second-level activations over arrangements of parts.

$$I(x) = \sum_{g\in G} \phi_0(g)a_0(g) + \epsilon(x), \quad a_0(g) = \sum_{k=1}^{K}\sum_{g\in G} \phi_1^k(g)a_1^k(g) + \epsilon(g) \tag{5}$$

Defining $\hat{I} = \sum_{g\in G} \phi_0(g)a_0(g)$ and $\hat{a}_0 = \sum_{k=1}^{K}\sum_{g\in G} \phi_1^k(g)a_1^k(g)$, the energy [2] and geometric LCA inference dynamics are given by

$$E = \frac{1}{2}||I - \hat{I}||_2^2 + \lambda_0 C(a_0) + \frac{1}{2}||a_0 - \hat{a}_0||_2^2 + \lambda_1 C(a_1)$$

$$\tau_0 \dot{u}_0(g) = -u_0(g) - [w_0 * a_0](g) + \langle \phi_0(g), I\rangle + \hat{a}_0(g)$$

$$\tau_1 \dot{u}_1^k(g) = -u_1^k(g) - [w_1^k * a_1^k](g) + a_1^k(g) + \langle \phi_1^k(g), a_0\rangle$$

4 Experiments

To evaluate and characterize the behavior of the proposed hierarchical, equivariant sparse coding model, we construct a synthetic dataset of scenes containing (1) objects comprised of lower-level parts where (2) parts and wholes are transformed via group actions. We do this by specifying a group G, constructing the dictionary elements at each level of the hierarchy, and then sampling from the generative model. For the first layer dictionary, we construct an overcomplete dictionary of Gabor functions generated by acting on a canonical Gabor template with a discrete sampling of the group of translations G. The mother Gabor ϕ_e is shown in Fig. 4. We construct $K = 2$ canonical second-layer dictionary elements ϕ_1^1, ϕ_1^2 from arrangements of parts at the preceding level of representation. Next, we generate the "orbit" of each template by again sampling from the group of translations G. The templates and selected dictionary elements are shown in Fig. 4. We then generate a dataset of images by sampling from the generative model. In particular, we create a sequence of frames in which objects present in the scene undergo different translations. The resulting images, inferred latents, and reconstructions are shown in Fig. 4. Note that the latent variables are *sparse* and *transform equivariantly*, as stated in the proofs.

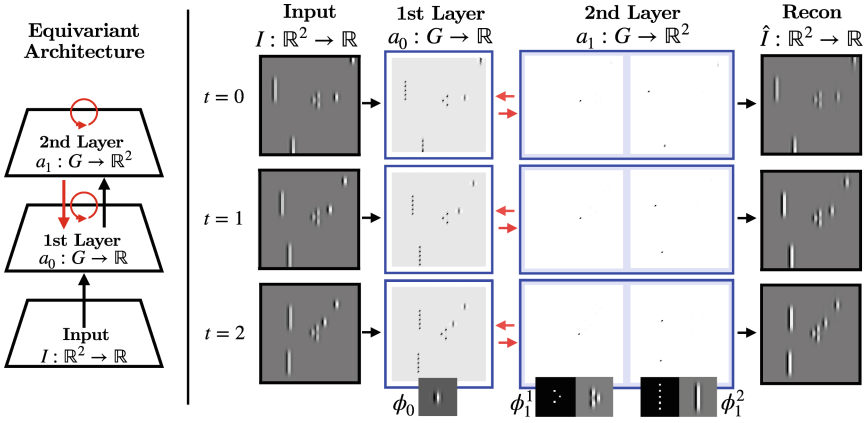

Fig. 4. Figure: (left) a two-layer translation-equivariant architecture with recurrent connections within and between layers, (right) experimental results demonstrating that the neural dynamics converge to a sparse, hierarchical representation of the scene which transforms equivariantly in time with the input video. Column 1: input video frames, Column 2: first layer gabor coefficient map displayed with sparse equivariant activations, Columns 3&4: two second layer "object" coefficient maps displayed with sparse equivariant activations

5 Discussion

By incorporating group structure, we have derived a new sparse coding model that is equivariant in its response to image transformations, both within a layer

and across multiple layers stacked in a hierarchy. We believe this is an important step toward developing a hierarchical, probabilistic model of visual cortex capable of performing perceptual inference (e.g. object recognition) on natural scenes. Surprisingly, the network architecture has the same functional form as the neural attractor model of Kechen Zhang [10], suggesting new circuit mechanisms in visual cortex for top-down steering, motion computation, and disparity estimation that could be done in the sparse code domain. Of relevance to deep learning, this new structure enables inference to be implemented efficiently on GPUs as (1) a feed-forward group convolution followed by (2) iterative lateral interaction dynamics implemented by group convolutions between dictionary elements.

Acknowledgements. The authors thank their helpful colleagues at the Redwood Center and Bioshape Lab. CS acknowledges support from the NIH NEI Training Grant T32EY007043.

References

1. Bekkers, E.J., Lafarge, M.W., Veta, M., Eppenhof, K.A.J., Pluim, J.P.W., Duits, R.: Roto-translation covariant convolutional networks for medical image analysis. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11070, pp. 440–448. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00928-1_50
2. Boutin, V., Franciosini, A., Ruffier, F., Perrinet, L.: Effect of top-down connections in hierarchical sparse coding. Neural Computation **32**(11), 2279–2309 (Nov 2020). https://doi.org/10.1162/neco_a_01325
3. Bronstein, M.M., Bruna, J., Cohen, T., Veličković, P.: Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. arXiv preprint arXiv:2104.13478 (2021)
4. Cohen, T., Welling, M.: Group equivariant convolutional networks. In: International Conference on Machine Learning, pp. 2990–2999. PMLR (2016)
5. Hall, B.C.: Lie groups, lie algebras, and representations. In: Quantum Theory for Mathematicians, pp. 333–366. Springer (2013). https://doi.org/10.1007/978-3-319-13467-3
6. Hyvärinen, A., Hurri, J., Väyrynen, J.: Bubbles: a unifying framework for low-level statistical properties of natural image sequences. JOSA **20**(7), 1237–1252 (2003). https://doi.org/10.1364/josaa.20.001237
7. Olshausen, B.A., Field, D.J.: Sparse coding with an overcomplete basis set: A strategy employed by v1? Vision. Res. **37**(23), 3311–3325 (1997)
8. Paiton, D.M., Shepard, S., Chan, K.H.R., Olshausen, B.A.: Subspace locally competitive algorithms. In: Proceedings of the Neuro-inspired Computational Elements Workshop, pp. 1–8 (2020)
9. Rozell, C.J., Johnson, D.H., Baraniuk, R.G., Olshausen, B.A.: Sparse coding via thresholding and local competition in neural circuits. Neural Comput. **20**(10), 2526–2563 (2008)
10. Zhang, K.: Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: A theory. J. Neurosci. **16**(6), 2112–2126 (1996)