The translation invariant bispectrum for feature analysis in complex cells

Sophia Sanborn [1] Ryan Zarcone [1] Bruno Olshausen [1] Christopher Hillar [1]

Abstract

We present a method for recovering the shape selectivity of complex cells (or other invariant neurons) via reverse correlation in the bispectral domain. Invariant representations of stimuli, such as the Fourier power spectrum, have been of great importance in the analysis of natural signals. However, the power spectrum has fundamental limitations, which we explicate here. We emphasize the need for a complete set of invariants, for which all points in a transformation orbit map to the same invariants, and distinct orbits map to distinct points. We advocate determining a complete sets of invariants with the machinery of the bispectrum, which can be effectively computed and inverted using the concrete representation theory of the underlying transformation group. As a proof-of-concept for the approach, we present a method for inferring invariant stimulus features for the special case of two-dimensional translation, demonstrating the potential of the 2D bispectrum for understanding the response properties of complex cells in visual cortex.

1. Introduction

Translation invariant features are critical components of stable representations of the visual world. This point was first noted theoretically in the works of (Poincaré, 1898) and (Pitts & McCulloch, 1947), observed empirically in visual cortex in the recordings of (Hubel & Wiesel, 1962), and most recently validated synthetically, with the success of convolutional neural networks at learning features that enable object recognition (Krizhevsky et al., 2012).

Complex cells are distinguished from simple cells by the complexity of the features they respond to and their invariance to slight shifts in the position of preferred features. The Fourier power spectrum has been used to capture some of these properties, e.g. (David & Gallant, 2005). However, complex cells show considerable sensitivity to the phase components of natural signals (Felsen et al., 2005), which cannot be captured by the phase invariant power spectrum. A translation invariant feature space that retains phase structure is thus highly desirable for the analysis and prediction of complex cell responses.

We further the exploration of invariants for the neural processing of natural signals by investigating the *bispectrum*, a general construction of features that do not change when data are acted upon by a transformation group. A direct application of the bispectrum concept first appears in work by statistician John Tukey to define higher-moments of time-series that can detect non-Gaussian phenomena (Tukey, 1953). We became aware of them from Kondor's applications to machine learning (Kondor, 2008), which, in turn, were inspired by papers of Kakarala (Kakarala, 1992). In fact, several researchers continually rediscover this circle of ideas (Gourd et al., 1989), even today (Cohen et al., 2018). More remarkably, however, already in 1947, considerations from neuroscience theory—predating evidence of "complex cells" from recordings in visual cortex—lead researchers to the idea of bispectra for canonical representations ("universality") in sensory coding (Pitts & McCulloch, 1947).

2. Bispectra for 2D Translation

Commonly, the Fourier power spectrum is used to determine features in stimuli that are invariant to translation. Although a powerful set of invariants for such purposes, the power spectrum is not a *complete* invariant—that is, a continuous map to a metric space with the following two properties: (1) all points in a transformation orbit map to the same invariants, and (2) distinct orbits are mapped differently. A simple demonstration of this fact can be seen in Figure 1, where it is shown that two very different stimuli have the exact same power spectrum. It follows that these features are insufficient to account for the ambiguities in stimuli up to 2D translation. On the other hand, bispectra form a complete representation up to the full transformation group.

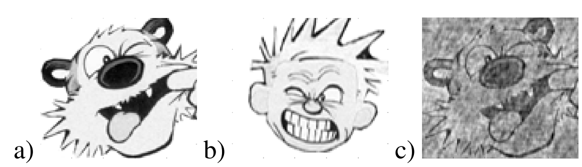

Figure 1. **Fourier power is an insufficient invariant**. a) Hobbes, b) Calvin, c) Calvin's 2D Fourier amplitudes with Hobbes' phases.

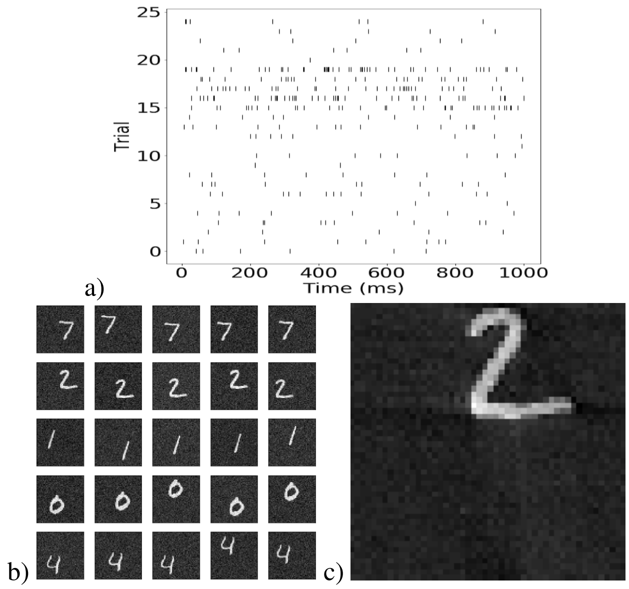

Figure 2. **Simulated translation invariant neuron**. a) Raster plot for a simulated Poisson neuron responding to presentations of MNIST digit stimuli. Simulated presentation occurred for 1000 ms. This example neuron was selective for "2"s. Thus, its firing rate significantly increased for trials 15 - 20, where 2s were presented. b) 25 MNIST digit stimuli. Each digit has been randomly translated and corrupted with Gaussian noise. c) Recovered stimulus from inverted reverse correlated bispectra.

In words, the 2D bispectrum is the Fourier transformation of the triple autocorrelation. Mathematically, this translates to the following formula. Given a 2D image $\mathbf{x} \in \mathbb{R}^{m \times n}$ and its corresponding Fourier transform $\mathbf{f} \in \mathbb{C}^{m \times n}$ ($\bar{\mathbf{f}}$ is the complex conjugate), the bispectrum $\beta \in \mathbb{C}^{m \times n \times m \times n}$ is:

$$\beta_{i,j,k,l}(\mathbf{x}) := \mathbf{f}_{i,j}\mathbf{f}_{k,l}\bar{\mathbf{f}}_{i+k \pmod m,\ j+l \pmod n}.$$

Although these bispectra form a complete set of invariants for 2D translation, only a small subset is required:

$$\beta_{1,1,1,1};\ \beta_{i,1,2,1},\ i = 1 \dots, m;$$

$$\beta_{1,j,1,2},\ j = 1 \dots, n;$$

$$\beta_{1,j,i,1},\ i = 2 \dots, m,\ j = 2 \dots, n.$$

In total, these number $mn + 2$, which is commensurate with the original data size and significantly easier to work with.

3. Methods & Results

We present a method for recovering the shape selectivity of complex cells via reverse correlation in the bispectral domain. As a proof of concept, we examine the performance of the bispectrum in decoding the responses of an artificial Poison neuron designed as a noisy translation invariant complex pattern detector (Figure 2). Each artificial neuron was presented fifteen noisy, translated MNIST digits from different classes. To generate a reconstructed preferred stimulus for each neuron, we invert the weighted mean of the bispectrum of each image presented to the neuron, with weights supplied by the neuron's responses to each image during the 1000 ms interval. Results for one neuron are shown in Figure 2, though reconstructions for all tested neurons show similar quality. Note that this method reproduces the neurons desired stimulus despite the presence of noisy neural responses to other digits and limited size dataset—and additionally removes the noise present in the original stimuli—suggesting a powerful tool for identifying translationally invariant neural representations.

References

Cohen, T., Geiger, M., Köhler, J., and Welling, M. Spherical CNNs. *arXiv preprint arXiv:1801.10130*, 2018.

David, S. and Gallant, J. Predicting neuronal responses during natural vision. *Network: Computation in Neural Systems*, 16(2-3):239–260, 2005.

Felsen, G., Touryan, J., Han, F., and Dan, Y. Cortical sensitivity to visual features in natural scenes. *PLoS biology*, 3(10):e342, 2005.

Gourd, F., Gauthier, J., and Younes, H. Une méthode d'invariants de l'analyse harmonique en reconnaissance de formes. *Traitement du signal*, 6(3):161–178, 1989.

Hubel, David H and Wiesel, Torsten N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology*, 160(1): 106–154, 1962.

Kakarala, R. *Triple corelation on groups*. PhD thesis, Department of Mathematics, UC Irvine, 1992.

Kondor, R. *Group theoretical methods in machine learning*. Columbia University, 2008.

Krizhevsky, Alex, Sutskever, Ilya, and Hinton, Geoffrey E. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pp. 1097–1105, 2012.

Pitts, W. and McCulloch, W. How we know universals: the perception of auditory and visual forms. *Bulletin of mathematical biophysics*, 9(3):127–147, 1947.

Poincaré, Henri. *On the Foundations of Geometry...* Open Court Publishing Company, 1898.

Tukey, J. The spectral representation and transformation properties of the higher moments of stationary time series. *Reprinted in The Collected Works of John W. Tukey*, 1: 165–184, 1953.