Analyzing the robustness of redundant population codes in sensory and feature extraction systems

Christopher J. Rozell * and Don H. Johnson

Department of Electrical & Computer Engineering, Rice University, Houston, TX, 77251

Abstract

Sensorineural systems often use groups of redundant neurons to represent stimulus information both during transduction and population coding of features. This redundancy makes the system more robust to corruption in the representation. We approximate neural coding as a projection of the stimulus onto a set of vectors, with the result encoded by spike trains. We use the formalism of frame theory to quantify the inherent noise reduction properties of such population codes. Additionally, computing features from the stimulus signal can also be thought of as projecting the coefficients of a sensory representation onto another set of vectors specific to the feature of interest. The conditions under which a combination of different features form a complete representation for the stimulus signal can be found through a recent extension to frame theory called "frames of subspaces." We extend the frame of subspaces theory to quantify the noise reduction properties of a collection of redundant feature spaces.

Key words: Feature extraction, population codes, redundant representations.

1 Introduction

Sensorineural systems often use groups of redundant neurons to represent stimulus information. This situation holds true in both the initial transduction of sensory information and the population coding of a single feature parameter. By using many neurons that overlap in the information they represent, a system becomes more robust to corruption (noise). We use the formalism of

* Corresponding author: *crozell@rice.edu*

frame theory to quantify the inherent noise reduction properties of such population codes. This linear vector space approach approximates neural coding as a projection of the stimulus onto a set of vectors, with the result (noisily) encoded by spike trains.

In addition, computing features from the stimulus signal can also be thought of as projecting the coefficients of a sensory representation onto another set of vectors specific to the feature of interest. The conditions under which a combination of different features form a complete representation for the stimulus signal can be found through a recent extension to frame theory called "frames of subspaces." We extend the frame of subspaces theory to quantify the noise reduction properties of a collection of redundant feature spaces.

2 Modeling sensory front-ends with frame theory

The initial receptors that transduce a stimulus signal in a particular sensory modality (e.g., audition, vision, etc.) can be approximated as a filtering operation over space $\vec{x}$ and time t. For instance, the response at time t of a unit with a receptive field at location $\vec{x}_k$ to the input signal s has the form of an inner product,

$$r_k(t) = \int h(\vec{x}, \vec{x}_k, \tau, t) s(\vec{x}, \tau)\, d\vec{x}\, d\tau, \tag{1}$$

where $h(\vec{x}, \vec{x}_k, \tau, t)$ is the response at time t of a receptor at location $\vec{x}_k$ to an impulse at time τ. Sensory front-ends provide complete information about all signals within their operating range (e.g., a space of spatially and temporally bandlimited signals). Because the sensing units generally overlap in bandwidth, amplitude range, and spatial coverage area, the representation of the scene by the collection $\{r_k(t)\}$ is *redundant*. This redundancy means that neural systems can reduce the communication and computational requirements by effectively transmitting lower resolution, noisy data and exploiting the redundancy to recover from errors.

Because of the limited temporal resolution achieved by biological systems, we will often assume for analytic reasons that the response is sampled at time steps t_m. Mathematically, the sensing operation described in equation (1) is a projection onto a set of vectors indexed by location and time sample. Expanding a signal in a set of orthogonal basis vectors is well-understood. However, because of the redundancy between the sensing units described earlier, the familiar analysis of an orthogonal basis does not apply. Instead, we rely on *frame theory* [1], which generalizes the notion of a basis to a redundant collection of vectors. The specific indexing of the receptive fields $\{r_k(t_m)\}$ by k and t_m isn't important to our mathematical results and the two-dimensional indexing can be cumbersome. For clarity in this report, we will simply discuss a collection of vectors $\{\phi_i\}$ indexed by i, which represents the receptive fields of a neural population covering a signal space $\mathcal{H}$ (which may be two-dimensional, such as the class of spatially and temporally bandlimited signals). This collection of vectors is called a *frame* for a Hilbert space $\mathcal{H}$ if there exist constants $0 < A \leq B < \infty$ such that for any signal $s \in \mathcal{H}$ the Parseval relation is bounded,

$$A \|s\|^2 \leq \sum_i |\langle s, \phi_i \rangle|^2 \leq B \|s\|^2 . \tag{2}$$

Typically, frame vectors are not linearly independent, meaning that *every* signal will normally have non-zero projections onto multiple frame vectors. When the frame vectors are normalized $\|\phi_i\|^2 = 1$, the frame is *uniform* and the constant A measures the frame's minimum redundancy. Frames with $A = B$ are called *tight*, and represent the case where the collection of sensing units preserve the energy in all signals uniformly well. A special case of a uniform tight frame is an orthonormal basis, which has $A = 1$.

Frame theory explicitly shows how to determine a set of *reconstruction* vectors $\{\widetilde{\phi}_i\}$ needed to reconstruct the signal from the coefficients $c_i = \langle s, \phi_i \rangle$,

$$s = \sum_i c_i \widetilde{\phi}_i . \tag{3}$$

While it is not clear that the brain would ever want to perform a reconstruction such as in equation (3), analyzing the reconstruction accuracy can serve as a proxy to assess how well the stimulus information is represented. One clear benefit of redundancy found in the stimulus representation is the reduction of corruption resulting from imperfect communication. Consider the case when the frame coefficients are corrupted by independent additive noise $n \sim N(0, \sigma^2)$ and the signal is reconstructed from the noisy coefficients

$$\widehat{s} = \sum_i (c_i + n) \widetilde{\phi}_i .$$

The frame's inherent redundancy reduces the mean-squared reconstruction error per signal dimension [1, 2] by

$$\frac{\sigma^2}{B} \leq \frac{\mathrm{E}[\|\widehat{s} - s\|^2]}{N} \leq \frac{\sigma^2}{A} . \tag{4}$$

This result means that by viewing the receptive fields described in equation (1) as a frame for some signal space, the lower frame bound A yields an upper bound on the corruption caused by imperfections due to additive noise.

While the robustness properties of a representation with regard to additive noise can be interesting, additive noise is a poor model for neural coding. Consider a population code that corresponds to a tight frame where the coefficients are encoded by a Poisson rate process, $\widetilde{c}_i = \text{Poisson}\left([c_i]_+\right)$, where $[\cdot]_+$ is a positive rectifier. We have calculated that reconstructing the signal with

Poisson encoded coefficients yields a MSE per signal dimension of

$$\frac{\mathrm{E}[\|\widehat{s}_p - s\|^2]}{N} = \frac{1}{A}\left(\frac{1}{M}\sum_{i=1}^{M} c_i\right),$$

where M the number of frame vectors. This result shows that the average noise variance (which is proportional to the coefficient magnitude) is also reduced by a factor of A in the Poisson encoding case.

3 Modeling feature extraction using frames of subspaces

Neural systems typically extract features from raw sensory input measurements. Even when a feature represents a single parameter of the signal, it is often encoded by a redundant population code [3]. If a feature is calculated by a linear operation (i.e., filtering), the feature calculation can be described as the projection of the measurement coefficients from primary receptive fields onto a set of redundant *feature* vectors. For example, the measurement coefficients corresponding to the receptive field for one sensing unit over several time samples could be projected onto a collection of feature vectors to perform spectral analysis. In the case of a Fourier transform calculation, no information about the stimulus is lost and the new coefficients can be thought of as also representing a frame for the signal space. However, many individual features calculated by sensory systems do not represent the full signal space (i.e., the resulting coefficients do not satisfy the lower bound in (2)). A population code for a feature that does not cover the whole input signal space represents a frame for a *subspace* of the signal space corresponding to that feature.

While one single feature may not represent the entire stimulus signal space, a collection of features computed in parallel might together form a complete representation. A formalism called "frames of subspaces" [4] provides methods for analyzing the relationships between the feature spaces. Mathematically, the ℓ^{th} feature space represents a redundant collection of vectors that span a subspace $W_\ell \subseteq \mathcal{H}$, with $\pi_\ell(s)$ denoting the projection of the signal s onto W_ℓ. The collection of L (possibly overlapping) subspaces $\{W_\ell\}$ is a *frame of subspaces* if, for any signal $s \in \mathcal{H}$, constants $0 < C \leq D < \infty$ exist such that

$$C\,\|s\|^2 \leq \sum_{\ell=1}^{L} \|\pi_\ell(s)\|^2 \leq D\,\|s\|^2. \tag{5}$$

A frame of subspaces is, in many ways, analogous to a frame. Feature spaces are not linearly independent and the bounds C and D measure the minimum and maximum redundancy (respectively) between the subspaces. Each feature space has its own redundant local representation formed by a population code that corresponds to a local frame (i.e., a frame only spanning that feature space) with bounds (A_ℓ, B_ℓ). Casazza [4] shows that the total collection of population codes representing all feature spaces is a frame for the signal space with frame bounds (A, B) satisfying $C \cdot A_{\max} \leq A$ and $B \leq D \cdot B_{\max}$, with equality when the same local frame bounds apply to each feature space. This result gives explicit conditions under which all of the feature-space representations together form a complete representation for the input signal space and it bounds the redundancy of that representation.

The noise reduction properties inherent in a single redundant population code are well known and have been discussed in section 2. In related work, we extended the frame of subspaces theory to quantify the noise reduction inherent in the redundancy between *subspaces* [5]. When feature spaces have corrupted population codes (i.e., perturbed by additive noise with a variance of σ^2) and an estimate $\widehat{s}_f$ is formed from reconstructed features, the MSE is bounded by

$$\frac{L\sigma^2}{B_{\max} D^2} \leq \frac{\mathrm{E}\|\widehat{s}_f - s\|}{N} \leq \frac{L\sigma^2}{A_{\min} C^2},$$

where L is the number of feature spaces and $(A_{\min}, B_{\max})$ are the extreme frame bounds of the local population codes. An interesting question to ask is how much noise reduction ability is lost by the system in performing feature-based processing (i.e., how much more could the noise be reduced if the population codes were all pooled and processed in a centralized way). Our results indicate that the upper bound on the noise reduction for a feature-based processing system is a factor of $\frac{L}{C}$ higher $(L \geq C)$ then a comparable centralized processing scheme. We have also shown that in the special case when the feature spaces cover the stimulus space equally $(C = D)$ and the underlying population codes representing each feature cover their respective feature spaces equally $(A_i = A_{\min} = B_i = B_{\max})$, the noise reduction abilities of a feature-based and a centralized processing scheme are equal (the system pays no penalty in noise reduction ability for processing each feature separately). While this mathematical condition seems restrictive, known asymptotic results [2] lead us to believe that as increasing numbers of features and population sizes tend to produce the necessary conditions.

4 Conclusions

Technological problems limit the amount of experimental data that can be collected from even a single functional population. It can be difficult to even determine the *size* of a functional population, let alone its information processing abilities. A theoretical understanding of the informational limits of both a single functional population and the interaction of several populations is necessary to help define a framework that yields experimental predictions and provides structure for analysis in the face of limited data. With the specific behavior of these populations unknown, it is necessary to start theoretical work with a general paradigm so that it can apply to *any* feature encoding population that may be found in the future. The formalism of frames and frames of subspaces begin to provide general tools for quantifying the properties of redundant populations.

While the analysis presented here is very abstract, we do hope to analyze specific known sensorineural systems more precisely in the future. As a concrete example, consider cells of the medial superior temporal area of the visual system which respond to at least three different types of stimulus parameters: expansion/contraction, rotation, and translation motions (left/right and up/down) [6]. The space of stimuli defined by these parameter combinations can be described as a four-dimensional space. Hypothetically speaking, 100 units with tuning curves uniformly distributed to respond to combinations of these stimuli features at a point in space would correspond to a tight frame able to reduce noise by a factor of $A = (100/4) = 25$. However, if the tuning curves were not uniform, A would be smaller. How much smaller depends the shape and distribution of the tuning curves. Thus, based on how the individual neuron's analyze information, we can derive the fundamental limits of accuracy of the population's representation. The sole caveat is the degree to which the representation process can be described as a linear projection.

References

[1] O. Christensen, An Introduction to Frames and Riesz Bases, Birkhauser, Boston, MA, 2002.

[2] V. Goyal, J. Kovačević, J. Kelner, Quantized frame expansions with erasures, Applied and Computational Harmonic Analysis 10 (2001) 203–233.

[3] F. Theunissen, J. Miller, Representations of sensory information in the cricket cercal sensory system. II. Information theoretic calculation of system accuracy and optimal tuning-curve widths of four primary interneurons., Journal of Neurophysiology 66 (1991) 1690–1703.

[4] P. Casazza, G. Kutyniok, Frames of subspaces, in: Wavelets, Frames and Operator Theory, Vol. 345, American Mathematical Society, 2004, pp. 87–113.

[5] C. Rozell, D. Johnson, Analysis of noise reduction in redundant expansions under distributed processing requirements, in: International Conference on Acoustics, Speech, and Signal Processing, Philadelphia, PA, 2005.

[6] M. Graziano, R. Anderson, R. Snowden, Tuning of MST neurons to spiral motions, The Journal of Neuroscience 14 (1) (1994) 54–67.