LINEARLY DECODABLE FUNCTIONS FROM NEURAL POPULATION CODES

M. B. WESTOVER, C. ELIASMITH, C.H. ANDERSON

ABSTRACT. The population vector is a linear *decoder* for an ensemble of neurons, whose response properties are nonlinear functions of the input vector. However, previous analyses of this decoder seem to have missed the obsevation that the population vector can also be used to estimate *functions* of the input vector. We explore how to use singular value decomposition to delineate the class of functions which are linearly decodable from a given population of noisy neural encoders.

1. INTRODUCTION

Many sensory systems utilize a large number of neurons to make measurements on inherently low dimensional systems. Examples include the $\sim$1000 neurons in the cricket cercal system that measure horizontal wind velocity [4]; and the much larger hair cell population in the mamallian otolith that senses linear acceleration of the head, a three dimensional vector [1]. Although the responses of the individual neurons are highly nonlinear and heterogeneous, a linear weighted sum of these nonlinear measures often provides an exceptionally precise linear estimate of the underlying physical input. Linear or "population vector" decoding is typically used to explore how neural systems might implement a communication channel; that is, how downstream populations might *reconstruct* the value of an encoded physical variable by specific synaptic weighting and summing of afferent sensory neuronal signals [6]. However, many downstream populations may be more concerned with extracting *transformations* of the sensory input. Using the general framework we present below, such transformations can also be computed by linear decoding.

The approach taken here is to view the computation of the decoding vecotr as a projection of the high dimensional state vector of neuronal activities onto a special set of axes. Traditional linear sensory input reconstruction is seen as a special case, involving a particular set of projections; combinations of projections along other axes provide measures of nonlinear functions. Examples include the magnitude and power, or square, of the input vector, or bilinear forms, like the vector cross product. The range of functions that can be extracted as well as their signal to noise ratios are discussed.

2. A GEOMETRIC VIEW OF LINEAR DECODING

Let $\mathbf{x}$ be a D-dimensional random vector representing the input to a set of N neurons, with response curves $\{a_1(\mathbf{x}), a_2(\mathbf{x}), \ldots, a_N(\mathbf{x})\}$. The response curves are

Date: February 12, 2001.

Key words and phrases. population codes, population vector, singular value decomposition,state space, principal component.

a heterogenous set of nonlinear, scalar-valued functions; and the input is physically restricted to a finite range R, e.g. $R = \{\mathbf{x} : \| \mathbf{x} \| \leq 1\}$. For simplicity, we assume that $\mathbf{x}$ is uniformly distributed on R.

We specify a set of N orthonormal vectors $\{\widehat{\mathbf{e}_1}, \widehat{\mathbf{e}_2}, \ldots, \widehat{\mathbf{e}_N}\}$ (columns of the NxN identity matrix) to serve as a basis (axes) for the population's state space. In these coordinates, the state vector for a given input $\mathbf{x}$ is

$$\mathbf{a}(\mathbf{x}) = a_1(\mathbf{x})\widehat{\mathbf{e}}_1 + a_2(\mathbf{x})\widehat{\mathbf{e}}_2 + \cdots + a_N(\mathbf{x})\widehat{\mathbf{e}}_N. \tag{2.1}$$

Thus, the neuronal population $\{a_i\}$ maps the range R onto a surface $\mathbf{a}(\mathbf{x})$ in the N-dimensional state space, parameterized by $\mathbf{x}$. Noise in the neural encoding process (e.g. jitter in spike times) introduces a sphere of uncertainty around each state point.

Next, let the desired transformation $f(\mathbf{x})$ be a vector in a generalized vector space $\mathbf{F}$ with inner product

$$< f_1(\mathbf{x}), f_2(\mathbf{x}) >= \int_R f_1(\mathbf{x}) f_2(\mathbf{x}) dx. \tag{2.2}$$

For example, we may take $\mathbf{F}$ to be the set of continuous functions over R.

Finally, we specify how the desired transformation is to be extracted from the neural activities by defining a *decoding rule:*

$$\widehat{f}(x) = \sum_{i=1}^{N} a_i(\mathbf{x})\phi_i = \mathbf{a}'(\mathbf{x})\phi, \tag{2.3}$$

where $\phi = [\phi_1, \phi_2, \ldots, \phi_N]'$ are the *decoding vectors*, or in the case of $D = 1$, *decoding weights.* Thus the extracted approximation to $f(\mathbf{x})$ will be a linear combination of the neural response functions. The set of all possible combinations defines a vector subspace $\widehat{\mathbf{F}} \subset \mathbf{F}$. We seek an optimal projector ϕ for projecting points along $\mathbf{a}(\mathbf{x})$ from the state space onto $\widehat{\mathbf{F}}$.

3. Problem statement and formal solution

Having set the stage, the precise statement and solution of our problem becomes geometrically obvious (see fig. 3.1a).

> *Find the function $\widehat{f}(\mathbf{x}) \in \widehat{\mathbf{F}}$ which best approximates $f(\mathbf{x}) \in \mathbf{F}$; that is, the vector in $\widehat{\mathbf{F}}$ which minimizes the norm of the approximation error $\epsilon = f(\mathbf{x}) - \widehat{f}(\mathbf{x})$.*

According to the well-known Projection Theorem, a unique solution $\widehat{f}_0(\mathbf{x})$ exists, and is simply the orthogonal projection of $f(\mathbf{x})$ onto $\widehat{\mathbf{F}}$. Thus, the error will be orthogonal to the subspace spanned by the neural response functions; i.e.,

$$0 \;=\; < f(\mathbf{x}) - \widehat{f}(\mathbf{x}), a_j(\mathbf{x}) >=< f(\mathbf{x}) - \sum_{i=1}^{N} a_i(\mathbf{x})\phi_i, a_j(\mathbf{x}) >, \tag{3.1}$$

for all $j = 1, \ldots, N$. Solving, we obtain

$$\phi_i = \sum_{j=1}^{N} \Gamma_{ij}^{-1} < f(\mathbf{x}), a_j(\mathbf{x}) > \tag{3.2}$$

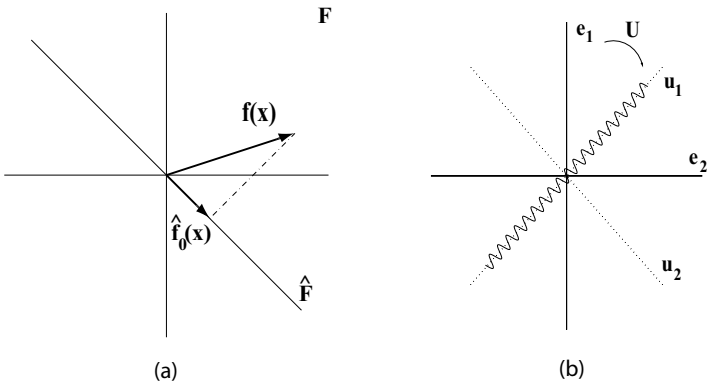

FIGURE 3.1. (a) Projection theorem Geometry. (b) Statespace rotation.

where

$$\Gamma_{ij} =< a_i(\mathbf{x}), a_j(\mathbf{x}) > \tag{3.3}$$

is the gram matrix, whose entries represent the correlations between activities of neuron pairs, averaged over the input range R.

4. DECOMPOSING THE NEURAL GRAM MATRIX

The gram matrix is usually ill-conditioned, so its inverse cannot be computed directly. This can be handled by adding statistically indepedent Gaussian noise to the neuronal responses [6]. Instead, we use the pseudoinverse based on singular value decomposition (SVD)[5, ch. 2], because it reveals much about what functions can be extracted from a population's activities.

Using SVD, we decompose the gram matrix as

$$\Gamma = \mathbf{U}'\mathbf{S}'\mathbf{U}, \tag{4.1}$$

where $\mathbf{S} = diag\{\omega_1, \omega_2, \ldots, \omega_N\}$ is the diagonal matrix of singular values in order of decreasing magnitude; and $\mathbf{U}$ is a *rotation* matrix whose columns are the corresponding eigenvectors. The columns $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_N\}$ of $\mathbf{U}$ define a new coordinate system for the state space.

In rotated coordinates, the state vector $\mathbf{a}(\mathbf{x})$ and the orthogonal projector ϕ become $\mathbf{U}\mathbf{a}(\mathbf{x}) \doteq \chi(\mathbf{x}) = [\chi_1(\mathbf{x}), \chi_2(\mathbf{x}), \ldots, \chi_N(\mathbf{x})]'$; and $\mathbf{U}\phi \doteq \Phi$. The function approximation can be reexpressed in rotated coordinates as

$$\widehat{f}(\mathbf{x}) = \mathbf{a}'(\mathbf{x})\phi = \mathbf{a}'(\mathbf{x})\mathbf{U}'\mathbf{U}\phi = [\mathbf{U}\mathbf{a}(\mathbf{x})]'[\mathbf{U}\phi] = \chi'(\mathbf{x})\Phi = \sum_{i=1}^{N} \chi_i(\mathbf{x})\Phi_i. \tag{4.2}$$

Each of the curves $\chi_i(\mathbf{x})$ represents the projection of the rotated state path $\chi(\mathbf{x})$ onto a coordinate axis $\mathbf{u}_i$. As before, $\widehat{f}(\mathbf{x})$ is a linear combination of functions spanning the subspace $\widehat{\mathbf{F}}$.

Crucial differences distinguish the new and old representations. First, in the original coordinate frame, the ordering of axes $\widehat{\mathbf{e}}_i$ is arbitrary, since there is no reason to prefer any particular response functions in the expansion of $f(\mathbf{x})$. However, projections along the rotated axes $\mathbf{u}_i$ *are* naturally ordered by the singular values of the gram matrix: Each singular value measures the second moment of the state

path $\chi(\mathbf{x})$ about its associated axis. Second, the spanning set $\{\chi_i(\mathbf{x})\}$ is now *orthogonal:*

$$< \chi_i(x), \chi_j(x) >= \omega_i \delta_{ij},$$

which means that the $\chi_i(\mathbf{x})$'s contribute *independent* information to the sum in (4.2), in decreasing order of importance from $i = 1$ to $i = N$.

Figure 3.1b illustrates these ideas for a hypothetical case where $D = 1$, $N = 2$. Before rotation, neither axis is indispensable to a representation of points on th curve $\mathbf{a}(x)$. After rotation, the projection along axis $\mathbf{u}_1$ clearly captures most of the information. Accordingly, truncating the expansion introduces only small changes in the error ϵ. In fact, noise in neuronal responses will destroy information along axes with small singular values, due to the cloud of uncertainty surrounding the deterministic state points [8].

By examining the first several χ functions (i.e., the robust projections of $\chi(\mathbf{x})$) we gain qualitative information about the nature and range of functions $\widehat{\mathbf{F}}$ supported by the population.

5. A SCALAR ENCODED BY MONOTONIC SATURATING REPONSE FUNCTIONS

We first examine a population with monotonic saturating functions that encode a scalar x (see fig. 5.1a). A well studied example of such is the population of neurons that encode horizontal eye position in the neural integrator [7]

In fig. 5.1b we plot the five $\chi_i(\mathbf{x})$ functions correponding to the highest singular values of the gram matrix. We see that this population supports an approximation basis which looks roughly like a "rotated" set of orthogonal polynomials over the interval $-1 \leq x \leq 1$. That is,

$$\{\chi_i(x) \approx \{P\{_i(x)\} = \{1, x, x^2 - 1/3, x^3 - 3x/5, ...\}.$$

As examples of what we mean by a rotation one can see that $\chi_0(x)$ is not a true constantas is $P_0(x)$ and $\chi_1(x)$ is not perfectly linear as is $P_1(x)$.

As expected, linear transformations $f : x \rightarrow \alpha x$ are well supported by the population, in addition to functions well approximated by low order polynomials (fig.5.1c). We emphasize *low* order, because projections onto later axes are small and thus sensitive to neuronal noise. For this population, we observe an approximately exponential decrease in magnitude of the singular values with increasing polynomial order (not shown). The number of $\chi_i(x)$'s with significant singular values, hence the range of transformations extractable from the population's neural activites, can be expanded by increasing the number of neurons or, perhaps more interestingly, by adjusting parameters governing the encoding of $\mathbf{x}$.

Note that this analysis fails if the nonlinear responses of the neurons in the population are *homogeneous.* Clearly, in this case, linear weighted sums will simply reflect the nonlinear responses of the individual neurons. This suggests that the diversity of real neuronal responses is computationally important.

6. THE SUBSPACE GENERATED BY GAUSSIAN TUNING CURVES

We now compare the above with a population of neurons having Gaussian tuning curves (see fig 5.1a), modeled after ensembles frequently encountered in mammalian cortical populations. Again we plot the first five projections of $\chi(x)$ along the eigenvectors of this population's gram matrix.

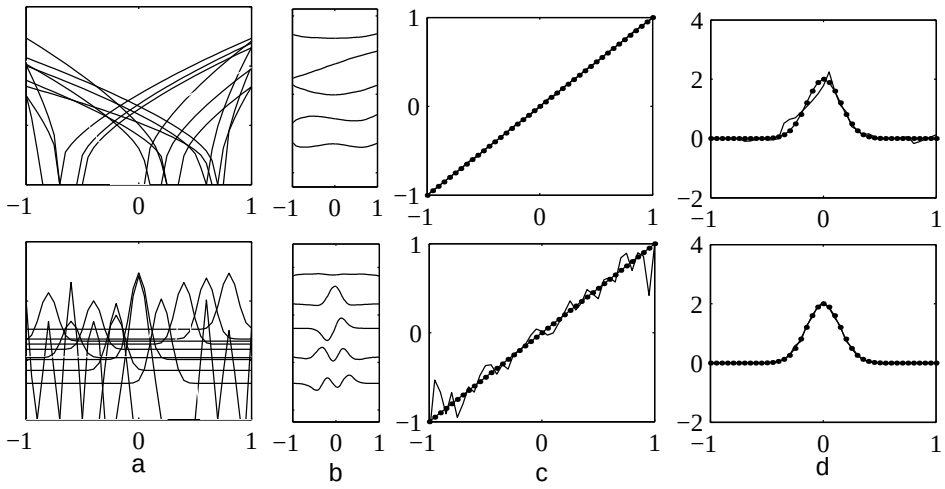

FIGURE 5.1. Top: Monotonic, broadly tuned population. Bottom: Gaussian-tuned population. (a) Tuning Curves. (b) First five $\chi_i(x)$'s. (c) Decoding results for $f(x) = x$. (d) Decoding results for a localized function $f(x)$. Dotted line is the desired function; solid line is the actual result. For (a),(c),and (d), N=20. For (b), N=200.

Several striking differences stand out. First, note the absence of a linear term. Predictably, attempts to extract linear transformations from this population are less successful (fig **5.1**c). However, the localized nature of these $\chi_i(x)$'s renders them a reasonable basis for approximating localized functions (5.1d).

Thus, through the lens of the rotated state space, the differences between the response functions of these populations are seen to imply qualitatively different computational potentialities. These results constrain hypotheses about the types of computations such populations may support.

7. CONCLUSION

We have described a method for defining the linear vector subspace generated by a population of neuronal response functions. For experimentalists, our method can facilitate exploration of the computations performed on incoming sensory signals by cortical areas. As we have shown previously, linear decoding also can facilitate direct computation of connection weights between populations in neural circuit models [2]. We have provided an examples illustrating that the class of linearly extractable functions closely depends on the heterogeneity and specific tuning properties of a population's neurons. The present treatment applies to the steady state; in future work we will extend our method to include the multitude of different temporal responses observed in real neuronal populations, such as those of the M and P retinal ganglion cells [9] .

REFERENCES

[1] D.E. Angelaki, M.Q. McHenry, J.D. Dickman, S. D. Newlands, and B.J.M. Hess, Computation of inertial motion: neural strategies to resolve ambigous otolith information, *Journal of Neuroscience 19* (1999) , 316-327.

[2] C. Eliasmith, and C.H. Anderson (forthcoming). *Neural Engineering: The principles of neurobiological sumulation.* Cambridge, MA: MIT Press.

[3] C. Eliasmith, and C.H. Anderson (in press). Beyond Bumps: Spiking networks that store sets of functions. *Neurocomputing*

[4] G.A. Jacobs, and F. E. Thenissen, Extraction of sensory parameters from a neural map by primary sensory interneurons. *Journal of Neuroscience 20* (2000) 2934-2943.

[5] W.H. Press; S. A. Teukolsky; W. T. Vetterling,; B. P. Flannery, *Numerical recipes in C* (Cambridge University Press; New York, NY, 1992).

[6] E. Salinas, and L.F. Abbott, Vector reconstruction from firing rates. *Journal of Computational Neuroscience 1* (1994) 89-107.

[7] H.S. Seung, D.D. Lee, B.Y. Reis, D.W. Tank, Stability of the memory of eye position in a recurrent network of conductane-based model neurons, *Neuron 26* (2000) *259-271.*

[8] C.E. Shannon, Communication in the presence of noise. *Proceedings of the IRE* 37(1) (1949) 10-21.

[9] D.C. Van Essen, and C.H. Anderson, Information processing strategies and pathways in the primate visual system. In Zornetzer S. F.; Davis, L. L; Lau, C.; McKenna, T (Eds.), *Neural and electronic networks.* (Academic Press, San Diego, CA, 1995).

[10] V.J. Wilson, and G.M. Jones, *Mammalian vestibular physiology.*(Plenum Press, New York, NY, 1979).