

Analysis of a Gain Control Model of V1: Is the Goal Redundancy Reduction?

Jianing Shi, Jim Wielaard, Paul Sajda
Department of Biomedical Engineering
Columbia University, New York, NY 10027

Abstract—In this paper we analyze a popular divisive normalization model of V1 with respect to the relationship between its underlying coding strategy and the extraclassical physiological responses of its constituent modeled neurons. Specifically we are interested in whether the optimization goal of redundancy reduction naturally leads to reasonable neural responses, including reasonable distributions of responses. The model is trained on an ensemble of natural images and tested using sinusoidal drifting gratings, with metrics such as suppression index and contrast dependent receptive field growth compared to the objective function values for a sample of neurons. We find that even though the divisive normalization model can produce “typical” neurons that agree with some neurophysiology data, distributions across samples do not agree with experimental data. Our results suggest that redundancy reduction itself is not necessarily causal of the observed extraclassical receptive field phenomena, and that additional optimization dimensions and/or biological constraints must be considered.

I. INTRODUCTION

The coding strategy of primary visual cortex (area V1) has been the subject of intense study for at least four decades. Recently, the question of coding by visual neurons has been considered with respect to natural viewing conditions, namely neural coding of natural images and sequences. Natural images exhibit higher-order statistical structure [1], [2] which should effect the encoding strategies used by neurons in V1. From an information theoretic point of view, it is natural to conjecture that V1 would encode natural image stimuli in such a way that optimizes its information representation capacity/efficiency under various biological constraints.

Parallel efforts in experimental neurophysiology and computational modeling have attempted to identify sensible mechanisms underlying an efficient V1 coding strategy. For example, Gallant and colleagues [3] showed, using natural movie stimuli, that the extraclassical receptive fields of neurons in V1 of macaque play an integral role for maximizing information efficiency of neurons and that this was consistent with a sparse coding hypothesis [4], [5]. Using computational modeling, Olshausen et al [5] used sparse coding methods and showed that structures resembling simple cell receptive

fields naturally fall out of such a coding strategy when applied to natural images.

A particular impressive demonstration of the consistency of a sensible coding strategy with observed experimental neurophysiology was provided by Schwartz and Simoncelli [1]. They argued for divisive gain control as a mechanism to reduce the redundancy between pairs of neurons in V1 and showed that after optimization over an objective function which reduces redundancy, several classical and extraclassical response phenomena observed experimentally in V1 neurons emerge, including contrast dependence, surround suppression, and receptive field growth at low contrast. Previous work by our group [6], [7] has focused on the neural circuitry underlying such classical and extraclassical responses and the diversity of responses seen experimentally and in realistic computational models. In contrast to the work by Schwartz and Simoncelli, we started with the neurophysiology and neuroanatomy and demonstrated circuitry which could produce not only typical cell responses but also realistic distributions of responses, in particular for surround suppression and contrast dependent receptive field growth. Since Schwartz and Simoncelli only reported “typical” cell responses we were interested in investigating if an objective to reduce redundancy in a divisive gain control model leads to realistic distributions of these extraclassical response properties and if the properties themselves are directly correlated to the minimization of the objective function.

II. METHODS

We constructed the divisive normalization model following the framework of [1], which includes a linear and non-linear stage together with an optimization for reducing redundancy between pairs of model neurons.

The linear response of V1 cells are approximated using a steerable pyramid [8]. The steerable pyramid is an overcomplete, multi-scale and multi-orientation decomposition with an impulse response similar to simple cell receptive fields of primary visual cortex [5], [9], [10]. The steerable pyramid enables one to represent different subbands, corresponding to V1 neurons with different spatial frequencies, positions and orientations.

As shown in [1] the joint distribution of subband coefficients from the steerable pyramid applied to an ensemble of natural images demonstrates higher order dependencies (dependencies beyond simple correlation). Specifically, if

This work was supported by NGA HM1582-05-C-0008

Jianing Shi, Jim Wielaard and Paul Sajda are with the Department of Biomedical Engineering, Columbia University, New York, NY 10027, USA
js2615@columbia.edu, djw21@columbia.edu and
ps629@columbia.edu

one takes pairs of subbands from the steerable pyramid, L_1 and L_2 , and looks at their conditional probability,

$$p(L_1|L_2) = \frac{p(L_1, L_2)}{p(L_2)}, \quad (1)$$

one sees variance dependencies, which visually resemble the shape of a “bowtie”. A similar analysis applied to white noise images does not demonstrate such a variance dependency.

Following the framework in [1] we assume that the marginal distributions $p(L_i|L_j)$ can be modeled as a Gaussian distribution.

$$p(L_i|L_j) = \frac{1}{\sqrt{\sum_j w_{ji} L_j^2 + \sigma_i^2}} \exp\left(-\frac{L_i^2}{2(\sum_j w_{ji} L_j^2 + \sigma_i^2)}\right). \quad (2)$$

We use maximum likelihood to fit the parameters of the model (w_{ji} and σ_i), where our data are the coefficients computed from applying the steerable pyramid to an ensemble of natural images.

$$\{w_{ji}, \sigma_i\} = \arg \max_{w_{ji}, \sigma_i} \left\{ \sum_x \log(p(L_i|L_j)) \right\} \quad (3)$$

Thus substituting Eqn. 2 into Eqn. 3 we obtain the objective function,

$$\{w_{ji}, \sigma_i\} = \arg \min_{w_{ji}, \sigma_i} \left\{ \mathbb{E} \left\{ \log(w_{ji} L_i^2 + \sigma_i^2) + \frac{L_i^2}{w_{ji} L_j^2 + \sigma_i^2} \right\} \right\}. \quad (4)$$

According to [1], redundancy reduction between a pair of neuron responses could be achieved using divisive normalization, by dividing out the weighted sum of rectified filter coefficients plus a constant,

$$R_i = \frac{L_i^2}{w_{ji} L_j^2 + \sigma_i^2} \quad (5)$$

where L_i is the filter coefficients of the steerable pyramid, R_i is the neuron responses after divisive normalization, w_{ji} and σ_i represent the weight and constant.

It is important to note that the optimization in Eqn. 4 makes several assumptions. First, for convenience, the marginal distribution $p(L_i|L_j)$, is parameterized as a Gaussian distribution. This assumption simplifies the calculation, yet ignores the fact that marginal distribution for the natural images is in general more peaked (higher kurtosis) than a Gaussian. A second assumption is that removal of redundancy in neural responses can be achieved by simply dividing out the “bowtie” variance dependence in a completely feed-forward manner, as opposed to using the recurrent network which the neurons are naturally embedded. This therefore means that the redundancy reduction is actually taking place between R_i and L_j , instead of R_i and R_j . Valerio and Navarro [11], propose a more rigorous approach by minimizing the mutual information between R_i and R_j , which results in a slightly different optimization and ultimately different weights and sigma. However for consistency we

use the approach outlined in [1] since this is in fact how comparison to the neurophysiology was made. The last major assumption or issue is due to the highly nonlinear form of the objective function, which results in many local-minima in the optimization. This is of particular interest to us since we were interested in investigating if all such minima lead to realistic physiological response properties.

To specifically consider this question, we ran the optimization over 300 times using an ensemble of natural images (standard images popular in image processing community, including baboon, einstein, lena, boat and etc. and also reportedly used in [1]). We solve for the optimal parameters numerically using conjugate gradient descent [12]. The weights, w_{ji} , are constrained to be positive to avoid numerical instability in optimization convergence [1]. Convergence of the optimization is reached when one of the six stopping criteria is reached (for details, see MATLAB help for `fmincon`). We save all the local minima as candidates for a sensible optimization result. Given the optimized parameters, we used drifting sinusoidal grating to investigate the response properties. There are three criteria we employed to judge whether a local minimum results in a reasonable solution: 1) orientation tuning, 2) size tuning and 3) contrast dependency.

All the neurons for all optimization runs showed reasonable orientation tuning and size tuning (surround suppression). However not all neurons showed realistic contrast dependence. Based on these results we divided our neurons into two groups: contrast-dependent (244 neurons who passed all three criteria) and contrast-independent (133 neurons who only passed the first two criteria).

For neurons that showed contrast dependence we further investigated the relationship between the value of the objective function and metrics for two extraclassical response properties (surround suppression and contrast-dependent receptive field growth). We looked at the distribution of receptive field size at both high contrast r_{f+} and low contrast r_{f-} , similarly surround size at high contrast r_{s+} and at low contrast r_{s-} . Receptive field size r_f is computed as the aperture size where the neural response reaches 99% of its maximum, while surround size r_s is computed as the aperture size where the neural response reaches 99% of its minimum.

The surround suppression index, which represents the degree of suppression, is given by Eqn. 6. f_{max} is the maximum of size tuning, f_{∞} is the neural response at the maximum aperture size in experiment, f_0 is the neural response at zero aperture size for stimulus.

$$SI = \frac{f_{max} - f_{\infty}}{f_{max} - f_0}. \quad (6)$$

The growth index, which describes the contrast dependency of neural response, is defined by the ratio of receptive field size / surround size at low contrast (0.06) and that of high contrast (0.25).

$$GI_f = \frac{r_{f-}}{r_{f+}}, GI_s = \frac{r_{s-}}{r_{s+}}, \quad (7)$$

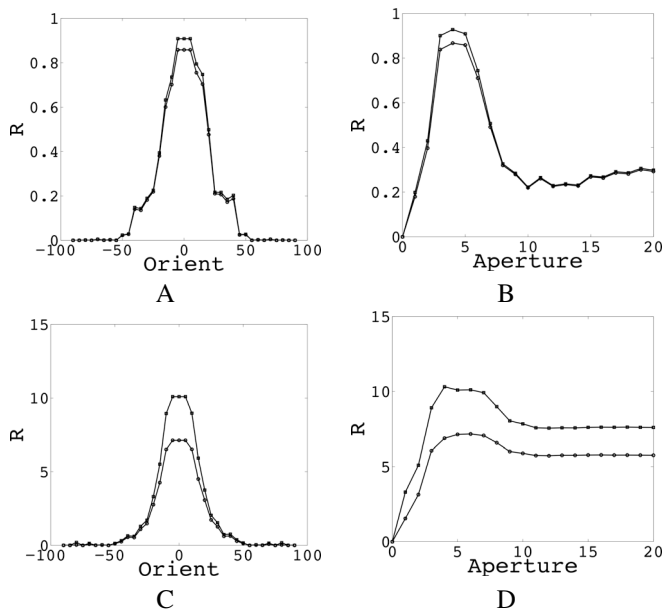


Fig. 1. A typical neuron, before and after optimization. This neuron shows orientation tuning (A) and surround suppression (B) before optimization, though the firing rate is very low compared with experimental data. After optimization, this neuron shows reasonable orientation tuning (C) and surround suppression (D). The low contrast receptive field growth (though very small in this example) emerges only after optimization. Squares are for high contrast stimuli, circles for low contrast.

where GI_f represents the growth index for receptive field size, while GI_s denotes the growth index for surround size.

III. RESULTS

Figure 1 shows responses of a neuron before and after optimization. Note that after optimization, the neuron shows orientation tuning, surround suppression and contrast-dependent receptive field growth. The orientation tuning and surround suppression are present even before optimization, though contrast-dependent receptive field size emerges only after optimization. This cell represents one of the cells that meets all three criteria described in the Methods section and thus is part of our further analysis (which we will refer to as our sample).

Figure 2 plots the value of the optimization function at convergence, for those cells satisfying all three criteria, against the two extraclassical response metrics we consider, growth index and suppression index. For the growth index we see three main states. For minima with larger values of the objective function we see no significant receptive field growth. At minima with lower values we see a transition to a state where growth is between 1.1 and 1.2. For local minima with the lowest objective function values, we see the greatest diversity, with some neurons with growth ratios up to 1.4, while others which in fact have growth ratios less than one (receptive field size reduces at low contrast). In terms of the suppression index, interesting is a systematic positive correlation between the suppression index and the value of the objective function. Thus for more optimal solutions (lower values of objective function) the surround suppression is negligible.

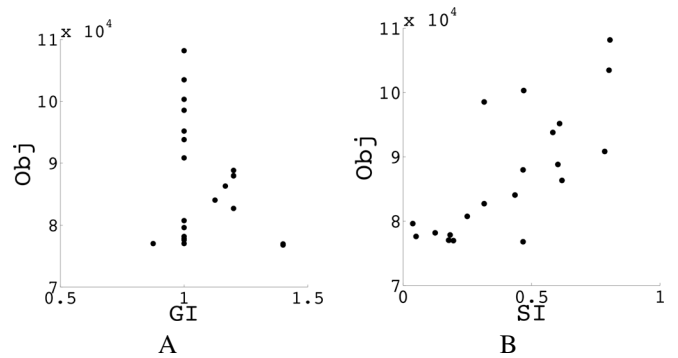


Fig. 2. (A) Objective function versus the growth index. The minimization of the objective function states the maximum independence among cortical cell responses. These are 21 neurons that meet all three criteria. Clearly, there is a zone with an objective function value in 82000–90000 that results in receptive field growth. (B) Objective function versus the surround index (SI). A strong correlation between the objective function and surround index is shown. The greater the objective function (meaning cortical neurons are less independent), the greater the degree of surround suppression.

Distributions of the receptive field size, surround size, together with suppression index and growth index, are shown in Figure 3. The distribution of suppression index is somewhat consistent with experimental data, with mean values close to what is seen experimentally (0.38 see [14]). However note that most of the distribution greater than 0.5 is due to local minima with objective function values significantly greater than 85000 (i.e. less redundancy reduction). Growth index values at low contrast are significantly less than what is seen experimentally and in other models, ([15], [6]). Finally we see a systematic dependence of the suppression index on contrast for our sample cells (see Figure 4), which is not consistent with experimental or other modeling data.

IV. DISCUSSION

The question of what is the information processing objective of early visual processing has been open since the the visual system became of interest for study. The focus of this paper has been to study whether a redundancy reduction objective is in fact consistent with experimentally observed physiological data. The nonlinear nature of the objective function creates multiple local minima in the optimization, which corresponds to diversity of the neural responses. Even though certain distributions for the population responses appear reasonable (surround index), details of the response properties are not consistent with experimental data. One important observation is that a neuron that agrees mostly with neurophysiological data does not correspond to the lowest objective function (meaning the most redundancy reduction). This suggests that there are other important optimization dimensions and/or biological constraints in the cortex coding strategy. It is possible that redundancy reduction is an attractive strategy in terms of information transmission, but well might not be the ultimate coding strategy for the cortex. Structures in the natural images, such as edge, might be distorted using a redundancy reduction coding strategy.

The model makes several assumptions that need to be further investigated. The Gaussian formulation for the marginal

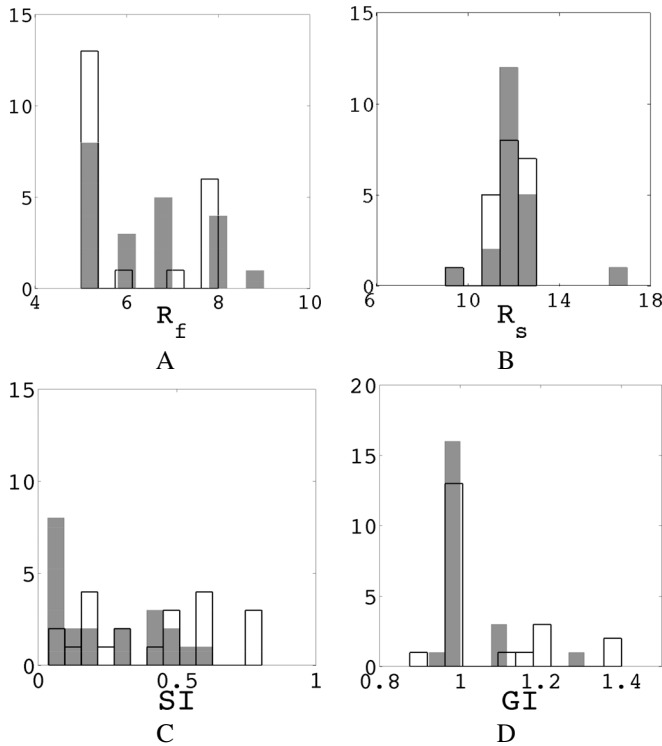


Fig. 3. (A) Distribution of receptive field size at low contrast (filled) with a mean of 6.38° , and at high contrast (white) with a mean of 6° . (B) Distribution of surround size at low contrast (filled) with a mean of 12.24° and at high contrast (white) with a mean of 11.95° . (C) Distribution of surround index at low contrast (filled) with a mean of 0.40, and at high contrast (white) with a mean of 0.25. (D) Distribution of growth index for receptive field size (white) with a mean of 1.07, and for surround size (filled) with a mean of 1.02.

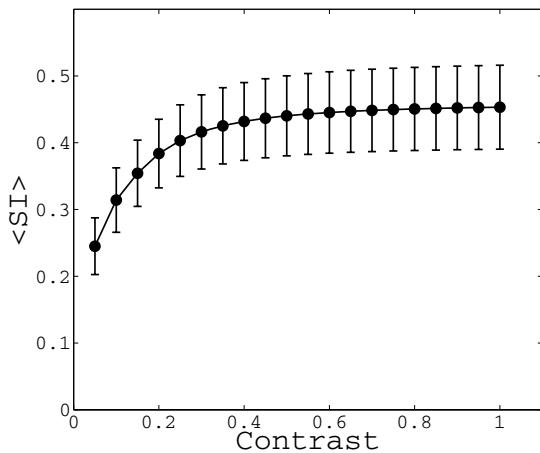


Fig. 4. Surround index as a function of contrast for the divisive normalization model. Clearly there is a very strong functional relationship between SI and contrast, which contradicts experimental data [13].

distribution is an approximate one, and only the “bowtie” dependency is considered in the model. The feedforward architecture also limits the dynamics of the model, let alone the redundancy removal is actually taken place between R_i and L_j , which might be an approximation that results in a different optimization direction than would be seen in “pure” redundancy reduction. Finally, the model only considers steady state visual responses, without time dependence. These assumptions might obscure other integral factors of an objective leading to an optimal coding strategy.

REFERENCES

- [1] Schwartz O. and Simoncelli E.P. Natural signal statistics and sensory gain control. *Nat. Neuroscience* 8, 819-825 (2001).
- [2] Simoncelli E.P. and Olshausen B. Natural image statistics and neural representation. *Annu. Rev. Neurosci.* 24, 1193-1216 (2001).
- [3] Vinje W.E. and Gallant J.L. Natural stimulation of the nonclassical receptive field increases information transmission efficiency in V1. *J. Neurosci.* 22(7), 2904-2915 (2002).
- [4] Barlow H.B. *Sensory Communication* MIT Press, Cambridge, Massachusetts (1961).
- [5] Olshausen B.A. and Field D.J. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* 381, 607-609 (1996).
- [6] Wielaard J. and Sajda P. Extraclassical receptive field phenomena & short-range connectivity in V1. *Cerebral Cortex* in press doi:10.1093/cercor/bhj090 (2005).
- [7] Wielaard J. and Sajda P. Neural mechanisms of contrast dependent receptive field size in V1. *Advances in Neural Information Processing Systems 18* 1505-1512. (2006).
- [8] Simoncelli E.P., Freeman W.T., Adelson E.H. and Heeger D.J. Shiftable multi-scale transforms. *IEEE Trans. Info. Theory* 38, 587-607 (1992).
- [9] Bell A.J. and Sejnowski T.J. The ‘independent components’ of natural scenes are edge filters. *Vision Res.* 37, 3327-3338 (1997).
- [10] van Hateren J.H. and van der Schaaf A. Independent component filters of natural images compared with simple cells in primary visual cortex. *Proc. R. Soc. Lond. B Biol. Sci.* 265, 359-366 (1998).
- [11] Valerio R. and Navarro R. Optimal coding through divisive normalization models of V1 neurons. *Comput. Neural Syst.* 14, 579-593 (2003).
- [12] Gill P.E., Murray W. and Wright M.H. *Practical Optimization* London, Academic Press (1981).
- [13] Sceniak M., Ringach D., Hawken M. and Shapley R. Contrast’s effect on spatial summation by macaque V1 neurons. *Nat Neurosci.* 2, 733-739, (1999).
- [14] Cavanaugh J.R., Bair W. and Movshon J.A. Nature and interaction of signals from the receptive field center and surround in macaque V1 neurons. *J. Neurophysiol.* 88, 2530-2546 (2002).
- [15] Cavanaugh J.R. *Properties of the Receptive Field Surround in Macaque Primary Visual Cortex*. Thesis, Center for Neural Science, New York Univ. (2000).