Sparsity, Matched Filters, and Natural Images

This section considers receptive field models from different perspectives. This includes the use of *sparsity* to suggest receptive field properties based on the statistics of natural images and also the idea of *matched filters* which revert to an older idea of receptive fields as feature detectors [99]. Sparsity was proposed by Barlow [7] as a general principle for modeling the brain based on the observation that typically only a small number of neurons are active. It was developed as a way to predict receptive field properties by Olshausen and Field [**?**]. It is natural to ask whether the receptive fields of cells encode basis functions which somehow capture the typical structure of images and represent it in a form which is suitable for later processing.

Our starting point is the idea that images, and particularly local regions of images. can be represented as a linear combination of basis functions $I(\vec{x}) = \sum_i \alpha_i \vec{b}_i(\vec{x})$, see equation (1).

Sparsity and Over-Complete Bases

Consider an image consisting of regions where the intensity varies spatially smoothly and regions where the intensity consists of a number of bright spots, or *impulses*. The smoothly varying regions of the image can be represented by fourier analysis efficiently, in the sense that we can approximate the intensity by only a small number of weighted sinusoids

By contrast, the impulses are much better represented in terms of a basis of impulse functions. It would be inefficient to represent them in terms of sinusoids.

In short, different types of basis functions are suitable for different regions of the image.

This suggests a strategy where we seek a representation in terms of an over-complete set of basis functions, in this case sinusoids and impulse functions, and a criterion which selects an efficient representation so that only a small number of basis functions are activated for each image. This requirement is called ℓ_1 *sparsity*.

ℓ_1 sparsity

More formally, we represent an image, or local image region, by:

$$I(\vec{x}) = \sum_{i=1}^{N} \alpha_i b_i(\vec{x}),$$

where the $\{b_i\}$ are the basis functions and the $\{\alpha_i\}$ are the coefficients. The number N of bases is bigger than the dimension of the image, and hence the bases are *over-complete*. Over-completeness implies that there are many ways to represent the image in terms of these basis functions (by different choices of the α's) and we need an additional criterion to select the α's. The ℓ_1 *sparsity* criterion proposes that we favor representations which make $\sum_{i=1}^{N} |\alpha_i|$ small, which penalize the weights of the basis functions and encourages most coefficients to be 0.

ℓ_1 sparsity criterion

We represent an image $\vec{I}$ by the approximation $\sum_{i=1}^{N} \hat{\alpha}_i \vec{b}_i$, where the $\{\hat{\alpha}_i\}$ are chosen to minimize the function:

$$E(\alpha) = \sum_{\vec{x}} (I(\vec{x}) - \sum_{i=1}^{N} \alpha_i b_i(\vec{x}))^2 + \lambda \sum_{i=1}^{N} |\alpha_i|. \tag{6}$$

The first penalizes the error of the approximation and the second term, whose strength is weighted by a parameter λ, penalizes the coefficients $\{\alpha_i\}$. The solution $\hat{\alpha} = \arg\min_{\alpha} E(\alpha)$ cannot be specified in closed form, but $E(\alpha)$ is a *convex* function of α and efficient algorithms exist for minimizing it to estimate $\hat{\alpha}$. The results of these algorithms can, for example, decompose an image into a sum of sinusoids and a sum of impulse functions.

Sparsity and Receptive Fields (I)

These ideas give an alternative way to think about the receptive fields of cells in V1. Firstly, observe that V1 has far more cells than the retina or the LGN and so it is has enough neural machinery to implement over-complete bases. Secondly, over-complete bases can be designed for specific image structures of interest (e.g., impulse functions or edges) which enables us to start interpreting the image instead of simply representing it. Thirdly, it relates to the observation that cells in V1 fire *sparsely*, which suggests [7] that they are tuned to specific stimuli and may relate to metabolic processes(firing a neuron takes energy which needs to be replenished). Hence the idea that the visual cortex seeks to obtain sparse, and hence presumably more easily interpretable representations, has intuitive appeal.

Sparsity and receptive fields (II)

Families of Gabor filters give an over-complete basis so they do not specify a unique representation of an image. These issues, and the relations of Gabors to wavelets, are discussed in more detail in [94].

Sparsity can be used to derive the properties of receptive fields of cell in V1 from natural images [127], see figure (18)(Left). Hence instead of hypothesizing models of receptive fields (e.g., Gabor filters) we can try to predict these receptive fields from studying images. These predictions do give some justification for Gabor functions but they also suggest other receptive field models which have also been experimentally observed.

Learning receptive fields using ℓ_1 sparsity

To learn the basis functions $\{\vec{b}_i\}$ from a set of natural images $\{\vec{I}^\mu : \mu \in \Lambda\}$ we extend equation (6) to obtain a criteria $E(b, \alpha)$ for fitting basis functions b and coefficients α to the set of images:

$$E(b, \alpha) = \sum_{\mu \in \Lambda} (I^\mu(\vec{x}) - \sum_{i=1}^{N} \alpha_i^\mu b_i(\vec{x}))^2 + \lambda \sum_{\mu \in \Lambda} \sum_{i=1}^{N} |\alpha_i|.$$

We estimate the basis functions $\hat{b}$ and the coefficients $\hat{\alpha}$ by minimizing $E(b, \alpha)$ to obtain:

$$(\hat{b}, \hat{\alpha}) = \arg \min_{(b,\alpha)} E(b, \alpha).$$

This criterion has been applied to natural images (where the $\vec{I}$ represent small image regions) and the resulting basis functions, see figure (18)(left), include filters which look like Gabor functions but they also include other types of filters which are also observed in experiments [127].

Alternatives: ICA

Other methods can predict receptive field properties from natural images using a similar image model, $I(\vec{x}) = \sum_{i=1}^{N} \alpha_i b_i(\vec{x})$, but imposing different assumptions on the form of the bases. In particular, independent component analysis (ICA) gives similar receptive field models [166]. Hyvarinen [67] explains this by showing that both types of models – L1 sparsity and ICA – both encourage the α_i to be strongly peaked at 0, but can occasionally have large non-zero values. If we remove the sparsity requirement and instead find the basis functions that minimize $\sum_{\mu \in \Lambda} (I^{\mu}(\vec{x}) - \sum_{i=1}^{N} \alpha_i^{\mu} b_i(\vec{x}))^2$? The basis functions will be the eigenvectors of the correlation matrix of the images and can be found by principal component analysis (PCA).

ℓ_1 sparsity figure

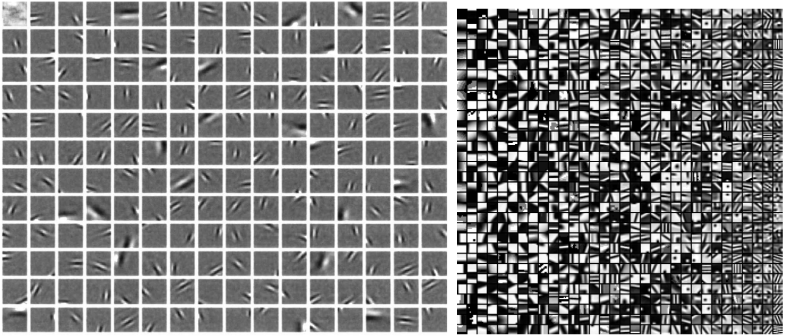

Figure 18: Left: The receptive fields learnt using sparsity [127]. Right: receptive fields learnt by matched filters.

Matched Filter Interpretation.

An alternative idea is that cells are feature detectors [99]. This can be modelled by a set of matched filters, which is an extreme form of sparsity, because any image patch can be represented by a single filter. Examples of matched filters are shown in figure (18)(right).

Suppose we have a filter $\vec{W}$ and an input image patch $\vec{I}_p$. We want to find the best fit of the filter to the image by allowing us to transform the filter by $\vec{W} \mapsto a\vec{W} + b\vec{e}$, where $\vec{e} = (1/\sqrt{N})(1, ..., 1)$. This corresponds to scaling the filter by a and adding a constant vector b. If $\vec{W}$ is a derivative filter then, by definition, $\vec{W} \cdot \vec{e} = 0$. We normalize $\vec{W}$ and $\vec{e}$ so that $\vec{W} \cdot \vec{W} = \vec{e} \cdot \vec{e} = 1$.

Matched Filters

The goal is to find the best scaling/contrast a and background b to minimize the match:

$$E(a, b) = |\vec{I_p} - a\vec{W} - b\vec{e}|^2.$$

The solution $\hat{a}, \hat{b}$ are given by (take derivatives of E with respect to a and b, recalling that $\vec{W}$ and $\vec{e}$ are normalized):

$$\hat{a} = \vec{W} \cdot I_p, \quad \hat{b} = \vec{e} \cdot \vec{I_p}.$$

The filter response is just the best estimate of the contrast a. The estimate of the background b is just the mean value of the image. Finally, the energy $E(\hat{a}, \hat{b})$ is a measure of how well the filter "matches" the input image.

Matched Filter Dictionaries

The idea of a matched filter leads naturally to the idea of having a"dictionary" of filters $\{\vec{W}^{\mu} : \mu \in \Lambda\}$, where different filters $\vec{W}^{\mu}$ are tuned to different types of image patches. In other words, the input image patch is encoded by the filter that best matches it. The dictionary of matched filters could be implemented by a set of cells (e.g., orientation columns). In this interpretation, the magnitude of the dot product $\vec{W} \cdot \vec{I}$ is less important than deciding which filter best matches the input $\vec{I}_p$. Matched filters can be thought of an extreme case of sparsity. In the previous sections, an image was represented by a linear combination of basis functions whose weights were penalizes by the ℓ_1, $\sum_i |\alpha_i|$. By comparison, matched filters represent an image by a single basis function. This gives an ever sparser representation of the image, but at the possible cost of a much larger image dictionary. Matched filters can be thought of as *feature detectors* because they respond only to very specific inputs.