Goals

Provide an overview of a major brain subsystem to help anchor concepts in neural network theory.
Discuss issues of representation.
Discuss functional requirements that determine the computations that networks must do.
Connect various parts and functions of the visual system with neural network ideas already studied, especially the (semi)linear networks, as well as ones coming up in remaining lectures of the course.

Neural networks and the visual system

Retina-cortex pathway

Roughly ten million retinal measurements are sent to the brain each second, where they are processed by some billion cortical neurons.

Retinal-cortical pathway: light->retina->thalamus<->cortex
light->(receptors->bipolar cells->ganglion cells)->(lateral geniculate cells)->V1 cells in layer 4C
...also other neurons (e.g. in the retina, horizontal and amacrine cells)
...also other pathways, e.g. to the superior colliculus
Summary approximation analogous to generic model neuron:

Despite several layers of synapses, many properties of ganglion cells, l.g.n. cells, and classes of V1 cells can be modeled using the generic model neuron, but with neural input replaced by light intensity. Where \( x_i \rightarrow L_i \), and \( L_i \) is the intensity of the light corresponding to location \( i \). So firing rate is:

\[
R_i = \sigma \left( \sum_{j=1}^{n} w_{ij} L_j \right)
\]

This corresponds to the steady-state lateral inhibition model we studied earlier, but has also been used to model cell responses in several areas of the mammalian visual system (e.g. "simple cells" of V1). We'll see below how more complex properties have been modeled using two layers of synaptic weights (e.g. "complex cells" of V1), corresponding to the non-linear multi-layer feedforward networks studied a couple of lectures ago.

We'll spend a considerable fraction of this lecture on the structure or spatial pattern of these weights.

There is an analogous model for time:

\[
R_T = \sigma \left( \sum_{t=1}^{T-1} w_{T,t} L_t \right)
\]

So what is new about the modeling? It's in the details. In vision we've learned a lot about the structure, extent, and possible development of the weights, and the forms of the squashing functions. We've also learned when the above generic model doesn't work. And also ideas about how to interpret the information processing functions of whole neural sub-populations modeled as non-linear feedforward networks in vision.
Pathways from eye-to-cortex

Schematic view

Retina

The primate retina has about $10^7$ cones that send visual signals to the optic nerve via about $10^6$ ganglion cells.
Ganglion cells do spatio-temporal filtering. They are band-pass filters meaning that low and high wiggles (measured as spatial frequency and temporal frequency) of image intensity get suppressed, and only a middle band of wiggles gets passed on. Why?

->Lateral inhibition as redundancy reduction.

Optic chiasm

The optic nerves from the two eyes meet at the optic chiasm where about half of the fibers cross over and the other half remain on the same side of the underside of the brain. Before synapsing in the lateral geniculate nucleus, about 20% of these fibers that now comprise the optic tract branch off to the superior colliculus--a structure involved with eye movements. The rest of the optic tract fibers synapse on cells in the lateral geniculate nucleus. Cells in the lateral geniculate nucleus send their axons in a bundle called the optic radiation to layer IV (one of six layers) of primary visual cortex.

Function: image sampling, contrast coding, gain control, spatio-temporal filtering (high pass--encoding change important), efficient coding

Functions of the Chiasm and Lateral geniculate nucleus (LGN)

The optic chiasm routes neuronal information so that information from corresponding points on the left and right eyes can come together at cortex for binocular vision, and in particular stereo vision. Typically animals with frontal vision have nearly complete cross-over, and animals with lateral eyes (e.g. fish) have little or no cross-over.

The nervous system has gone to considerable length to bring information from the two eyes together early on. Although there seems to be little if any binocular interaction between neurons in the LGN, the arrangement of the optic chiasm is the first step towards the eventual construction of a topographic cortical map with information from both eyes.

In fact, there is a general principle that becomes even more apparent when one looks at maps that pervade cortical organization:

Neural computations that require bringing information together often require close physical connectivity between neurons

Anderson has a discussion of topographic and tonotopic maps in cortex. Later on we will see some of the consequences of self-organizing principles that serve to minimize wiring length when we study Kohonen networks and adaptive maps.

The neurons of lateral geniculate nucleus do more band-pass filtering, and the cells are characterized by fairly symmetrical center-surround organization like the ganglion cells. They show even less response to uniform illumination than ganglion cells. Despite the fact that neurons from the two eyes exist within the same nucleus, no binocular neurons are found in LGN. We have to wait until cortex to see binocular neurons. Although the LGN is often considered a relay station, feedback from cortex suggests possible role of attention mechanisms (see Crick, 1984 for a speculative neural network theory of LGN and reticular function). Sillito et al. (1994) have found "Feature-linked synchronization of thalamic relay cell firing induced by feedback from the visual cortex". Nature, 369, N. 9, 479-482. See Sherman and Guillery (2002) for a recent discussion of thalamo-cortical anatomy.

The superior colliculus has a key role in the control of eye movements--a highly non-trivial control problem requiring coordination of head and eye movements in the context of a constantly changing environment.
Primary Visual Cortex

Cortex in general

Recall: four lobes: occipital, temporal, parietal, frontal.

~14 inches in diameter if flattened, ~$10^{10}$ neurons. 2000 to 10000 synapses each.

Cortical modularity:

  e.g. ~30 visual areas, ~12 with some degree of topography.

Largely inhibitory (GABA), typically quiet (metabolic cost to firing, Lennie (2003)). Excitatory transmission (glutamate).

(Lots of other transmitters).

Long range excitatory pyramidal cells (between areas). Within area excitatory and inhibitory connections.

Primary visual cortex: Large scale organization

- Overview

  primary visual cortex (striate cortex, V1 in primates, area 17 in cats)
  anatomical organization - topographic, later lecture on adaptative maps
  functional cell types: simple, complex, end-stopped
  model of simple cells

    *generic feedforward neural network models*

    other than sigmoid non-linearities?
**Topographic map**

The striate cortex is laid out as non-linear topographic map with 80% of cortical area devoted to about 20% of visual field, reflecting the higher acuity of foveal vision. Because of the cross-over at the optic chiasm, the left visual field (right retina) maps to right hemisphere. (see Engel et al., 1997 for fMRI measurements of human cortical magnification). Nearby locations in the retina map to nearby locations in primary visual cortex--thus, V1 is said to have a retinotopic map.

There are a number of interesting features of this map, where locations in retinal polar coordinates are approximately mapped to rectangular coordinates on V1. Lecture 21 describes the properties of this map in more detail.

Why is visual information laid out topographically in many cortical areas?

**Layers**

Axons from the optic radiation synapse on layer IV neurons of the primary visual cortex (area 17, striate cortex or V1). Cortex is anatomically structured in layers, numbered from I (superficial) to VI (deep). Connectivity between areas is closely tied to layer structure.

**Primary visual cortex: Neuron properties**

V1 neurons respond well to contrast, both in terms of absolute sensitivity and contrast differences (Geisler and Albrecht, 1997; Boynton et al., 1999; Olman et al., 2004). Apart from the neurons the LGN fibers synapse on, and in contrast with receptive field characteristics of earlier neurons, many cortical cells are tuned or selective to local stimulus attributes.

Tuning varies along the dimensions of:

- orientation (θ)
- degree of binocularity
- motion direction
- spatial frequency (f, or fx, fy)
- spatial phase (ϕ)

Some of the major contributions to our understanding of visual cortex was due to the research of Hubel, D. H., & Wiesel, T. N. (see 1968 reference). Hubel and Wiesel won the Nobel prize for their work.

**Columns and hypercolumns: orientation and degree of binocularity**

The cells of the primary cortex are organized into columns running roughly perpendicular to the surface in which cells tend to have the same orientation preference and degree of binocularity. A "hypercolumn" is a group of columns spanning all orientations. The size is on the order of cubic millimeters and includes about 100,000 cells. Different hypercolumns have different receptive field centers. In the cortex, we see for the first time binocular cells.

Embedded in the cortical hypercolumns are cytochrome oxidase blobs in which are found opponent color cells that seem to lack strong orientation selectivity (Livingstone, M. S., & Hubel, D. H., 1984; Livingstone, M. S., & Hubel, D. H., 1987).
29–10. This basic cortical module (hypercolumn) in area 17 of the visual cortex contains a complete set of architectonic columns representing 500° and a set of ocular dominance columns. Each hypercolumn also contains several cortical puffs, regions of cortex to which the cells do not have an area of termination. The cells in the puffs are connected with color, as we shall see in Chapter 30. Each layer of the lateral geniculate nucleus receives input from either the contralateral (C) or the ipsilateral (I) eye and projects in turn to the ipsilateral or the contralateral ocular dominance columns.

29–7. Comparison of the receptive fields of neurons in the primary and the lateral geniculate nuclei with those of extrastriate cortex in area 17. A. Cells of the primary and lateral geniculate nuclei fall into two classes: center (I) and surround (II). B. Neurons of the primary visual cortex fall into two classes: simple and complex. Each of these classes, moreover, has several subtypes. C. Thus, there are at least three subtypes of neurons: simple, complex, and interneurons. D–F. Specific maps of orientation. In contrast, only receptive fields with a vertical axis of orientation from 12 to 6 o'clock are shown in the figures. Each has a rectangular configuration. In fact, each region of the retina is represented in area 17, not only the area but all sizes of orientation: vertical, horizontal, and various oblique (adapted from Hubel and Wiesel, 1962).

28–14. The afferent and efferent connections of the primary visual cortex are made in specific layers of cortex.

If one views the cortex as essentially a 2-D sheet, it is reasonable to ask how a high-dimensional feature space can be mapped into it and maintain a requirement that similar features (e.g. orientation, ocularity, spatial frequency, motion selectivity) get mapped to near-by locations. We return to this problem in the lecture on adaptive maps.

■ Simple cells

There are two main types of cells. The simple cells are roughly linear except for rectification, are spatially and temporally band-pass, and show spatial phase sensitivity. A first approximation model for simple cell response firing rate (in impulses/sec) is:

\[ R = \sigma \left[ \sum_{i,j} W_{ij} L_{ij} \right] \]

Where \( W_{ij} \) are the receptive field weights, and \( L_{ij} \) the image intensity values at spatial location \((i,j)\). An example would be a Gabor function (see discussion and figures below):

\[ W_{ij} = e^{\frac{-i^2-j^2}{2\sigma^2}} \cos(2\pi(f_x i + f_y j) + \phi) \]

We can visualize the weights for particular values of the horizontal and vertical spatial frequencies \( f_x, f_y \), and phase \( \phi \):

The squashing function is a half-wave rectification operation, \( \sigma \) sets negative values to zero, and is linear for positive values:
You can see that this has the same form as the **generic neuron model**, except that the inputs are the physical stimulus values.

And as we saw at the beginning of the course, a better model is obtained by replacing the straight sloping line with one that saturates at high values. This model is steady state. To include time domain dependencies requires the introduction of a band-pass temporal tuning characteristics.

### Complex cells

The second major class of neurons is that of **complex** cells. Like simple cells, complex cells are spatially and temporally band-pass, show orientation and motion direction selectivity, but are insensitive to the phase of a stimulus such as a sine-wave grating. Rather than half-wave rectification, they show full-wave rectification. A model for complex cells would resemble the sum of the outputs of several subunits positioned at several nearby spatial locations. Each subunit would resemble simple cell with a linear spatial filter followed by a threshold non-linearity.

If true, the simplest neural-net like version of this model would correspond to two layers of weights, where the first set feed into simple cells, and the second set feed into complex cells. In actuality, complex cells may not be built out of simple cells, and as mentioned above, the generic connections model of simple cells collapses a number of neural layers to one effective layer. Another complication is that cells show a property called "response normalization" (see contrast normalization, below).

One way of obtaining the phase insensitivity would be to use subunits with cosine and sine phase receptive fields. We see below how a neural network can be built that can be used to detect edges—it combines simple cell outputs into outputs similar to those of complex cells.

The motion selectivity could be built in with appropriate inhibitory connections between subunits. Full-wave rectification could be built with subunit pairs that have excitatory and inhibitory receptive fields centers.

### Contrast normalization

Both simple and complex cells show contrast normalization—a feature not included in the above simple model. For a discussion of steady-state models of simple and complex cells see Heeger (1991) and Carandini et al. (1997)
\[ R_i = \sigma \left( \sum_{j=1}^{n} w_{ij} L_j \right) / \sum_{k=0}^{R_k^2} \]

where \( N_i \) is a neighborhood of neuron \( i \)

- **End-stopped cells**

A third class of cells are the end-stopped (or "hyper-complex") cells that have an optimal orientation for a bar or edge stimulus, but fire most actively if the bar or edge terminates within the receptive field, rather than extending beyond it. It has been suggested that these cells act as "curvature" detectors. (Dobbins, A., Zucker, S. W., & Cynader, M. S., 1987).

These cells are also thought to be important for detecting occluding surfaces and the perception of illusory contours. (Heiter et al., 1992).

Whether or not these end-stopped cells should be considered a distinct functional class has been a matter of debate.

### Functions of Primary Cortex

- **Functions of primary visual cortex**

  Emphasize local image measurements likely to be caused by surface properties, e.g. abrupt intensity changes due to object edges. But correlation is weak.

- **Binocular vision and Stereo**

  As mentioned earlier, primary cortex brings together information from the two eyes in single neurons. This information is important for coordinated eye movements and stereo vision. Although V1 cells are predominantly binocular, it was at first thought that disparity selectivity did not arise until V2 (Hubel and Wiesel, 1970). However, there is evidence for disparity selective cells in V1 and V2 (Poggio, G., F., & Poggio, T., 1984). Disparity selectivity is a trivial task for single bar stimuli, and later neurons were found that effectively solve the problem of false matching (Poggio and Talbot, 1981).


  This algorithm is related to **Hopfield networks** that we will study later in this course.

  Stereo vision has received a lot of attention in both computer and biological vision over the last 15 years. Later we will look at a neural network model of stereopsis.

- **Motion**

  motion measurements--local in space and time

  The directional selectivity of cells in striate cortex provide a form of early motion detection, akin to that described for invertebrate and rabbit peripheral vision. This detection is only local and thus ambiguous. Cortical cells suffer from the
"aperture problem", and further computation is required to disambiguate object motion. Cortical cells are also selective for speed (Orban et al., 1983). Both the motion selectivity and binocularity suggest a general hypothesis for cortical function: it links information likely to have a single environmental cause for subsequent extra-striate processing.

Anderson discusses a neural network solution to the aperture problem in Chapter 10.

**Spatio-temporal filtering**

Overview

Spatial frequency filtering, temporal filtering - *generic feedforward neural network models* (Lecture 3)

Why spatial filtering?

- edge detection

  cortical basis set and economic representations - *self-organizing neural networks (Lecture 14)*

Why temporal filtering?

- transient detection

  economic representations

Predicative coding? - A key theme in information processing by the brain

Beginning with the psychophysical results of Campbell and Robson (1968), and continuing with studies of the spatial and temporal frequency selectivity of simple and complex cells, there has emerged a detailed picture of how images may be processed in the visual cortex. Let us look at spatial frequency in detail with a view to understanding its computational function in vision.

**Spatial filtering**

**The convolution model for neural networks**

Spatial filtering can be realized as a linear feedforward network. Point-wise non-linearities, and response normalization need to be added on to improve models’ accounts of the data.

Recall our model of lateral inhibition. There are both ON-center and OFF-center types of ganglion cells, specified by the weights. We noted that we could quantitatively model an array of ganglion cell responses as a matrix operation acting on an input image vector.

We didn't make a deal of it at the time, but over small regions of space, neural image processing can be approximated as homogeneous. Thus we assumed "shift-invariance" (outputs are shifted linearly with shifts in inputs) which implies a particular structure and properties to the weight matrix—i.e. in 1D, the rows are identical except for progressive shifts as we go from one row to the next. Thus the rows the weight matrix are just shifted versions of each other. In other words, the basic properties of spatial filtering are reflected in the weight structure of a single neuron. The receptive fields for other neurons are basically the same, except for position. From a mathematical point of view, if space is treated as continuous, such homogeneous linear spatial filtering is called "convolution".
Let \( r_{k,l} \) be the response (in spikes/sec) of a ganglion cell at x-y location \((k,l)\). The average response, to a first approximation, is determined by the weighted sum of the inputs, \( g_{i,j} \) at spatial location \((i,j)\)

\[
r_{k,l} = \sum_{i,j} w_{k,l; i,j} g_{i,j}
\]

If we assume spatial homogeneity, and thus shift-invariance:

\[
r_{k,l} = \sum_{i,j} w_{k-i; l-j} g_{i,j}.
\]

Or by suitable arrangement of rows and columns as matrix operation, \( r = W.g \)

In the continuous case, by the convolution integral:

\[
r(x,y) = w^* g = \int \int w(x - x', y - y') g(x', y) \, dx' \, dy'.
\]

The fundamental computations in spatial filtering of input images are linear--image vectors are multiplied by a weight matrix.

When it reached steady-state, the lateral inhibition network studied in Lecture 3 was essentially convolving the input with the weights. If you ever use a graphics package like Adobe Photoshop, you can easily convolve the image on the computer screen with any number of possible spatial filters (i.e. weight matrices). *Mathematica* has a built-in function `ListConvolve[ ]` that accepts as arguments an input vector and a "kernel". But we need to know how to specify the "kernel", i.e. the weights for the convolution.

So the questions are:

Can other neural systems be modeled as taking convolutions of their inputs? Yes. Simple cell responses are modeled as convolutions with a point-wise non-linearity.

What is the structure of the effective weights for simple cell neurons in the visual cortex? Let's look at this more closely because it provides insight into issues of neural representation.

### Basis set for representing visual information

*Psychophysics and physiology*

The results of masking, adaptation, and other psychophysical studies of spatial and orientation frequency selectivity in human vision are surprisingly consistent.

What is the form of the weights \( w_{ij} = w_{xy} \)?

*A cortical basis set for images specifies the effective weights as a function of spatial position*

Both the psychophysical and neurophysiological data could be accounted for, in part, by assuming the visual system performed a quasi-Fourier analysis of the image. One possible model assumes that the visual system computes the coefficients (or spectrum) of an image with respect to the following basis set, called a Gabor set (Daugman, 1988):

\[
\begin{vmatrix}
 k e^{-\frac{(x^2 + y^2)}{2a^2}} \cos(2\pi(f_x x + f_y y)), \\
 k e^{-\frac{(x^2 + y^2)}{2a^2}} \sin(2\pi(f_x x + f_y y))
\end{vmatrix}
\]

The spectrum coefficients are represented by the firing rates of cells whose receptive field weights are represented by the above basis functions. In actuality, because as we saw earlier, simple cells behave more like linear filters followed by half-wave rectification, there should be two cells for each coefficient--"on" and "off" cells). One difference between this
The spectrum coefficients are represented by the firing rates of cells whose receptive field weights are represented by the above basis functions. In actuality, because as we saw earlier, simple cells behave more like linear filters followed by half-wave rectification, there should be two cells for each coefficient—"on" and "off" cells. One difference between this basis set, and the Fourier basis set (i.e. the optical eigenfunctions) is that this set has a local spatial restriction because of the Gaussian envelope. A second difference, which has major implications for computation, is that the basis functions are, in general, not orthogonal. Graphs of these functions typically look like wave-packets:

Here we've plotted a one-dimensional slice (e.g., \( w_x \)) through a sine, and cosine Gabor function. In two dimensions (with the standard deviation, and the x and y spatial frequencies equal to 1), we can visualize the receptive field weights as follows.

■ **Visualizing the Gabor functions:**

```
 cgabor[x_,y_, fx_, fy_, s_] :=  
 Exp[-(x^2 + y^2)/s^2] Cos[2 Pi(fx x + fy y)];  
 sgabor[x_,y_, fx_, fy_, s_] :=  
 Exp[-(x^2 + y^2)/s^2] Sin[2 Pi(fx x + fy y)];
```

■ **Relation of horizontal and vertical frequencies to orientation and oriented frequency**

\[
\text{orientation} : \quad \theta = \tan^{-1}\left(\frac{f_y}{f_x}\right)
\]

\[
\text{frequency} : \quad f = \sqrt{f_x^2 + f_y^2}
\]
Make a list of the orientations, and center frequencies of the basis set

\[
\text{vtheta} = \text{Table}[i \pi/4, \{i, 4\}]; \\
vf = \{0.25, 1, 4, 8\};
\]
Plot the elementary basis functions with the width, $s$, proportional to the reciprocal of spatial frequency. This maintains a constant bandwidth in octaves.

We’ve discretized the above continuous specification of the basis set. This leaves several free parameters. Most models of detection and masking get by with about no more than 6 spatial frequencies, about 12 orientations (specified by the ratio of horizontal and vertical spatial frequencies), and two phases (cosine and sine) at each retinal location. A subset of neurons representing a particular spatial frequency bandwidth makes up a spatial frequency channel. (Although there is neurophysiological evidence for pairs of V1 neurons having receptive fields with 90 deg phase shifted relative to each other, there is evidence against absolute phase--i.e. there is not a predominance of edge or bar type receptive fields. See Field and
We've discretized the above continuous specification of the basis set. This leaves several free parameters. Most models of detection and masking get by with about 6 spatial frequencies, about 12 orientations (specified by the ratio of horizontal and vertical spatial frequencies), and two phases (cosine and sine) at each retinal location. A subset of neurons representing a particular spatial frequency bandwidth makes up a spatial frequency channel. (Although there is neurophysiological evidence for pairs of V1 neurons having receptive fields with 90 deg phase shifted relative to each other, there is evidence against absolute phase—i.e. there is not a predominance of edge or bar type receptive fields. See Field and Tolhurst).

One parameter still left unspecified is the standard deviation or spread of the Gaussian envelope. If large, this basis set approaches that of regular and global Fourier analysis. The psychophysical data suggest that the standard deviation be such that the Gaussian envelope is about one cycle (at the 1/e point) of the sine wave. One cycle corresponds to about 1.5 octaves spatial frequency bandwidth.

Why a local Gabor-function representation of visual information?

Why would the visual system have such a representation? We have two types of explanations. One is that encoding over multiple spatial scales is important for subsequent processing that may involve edge detection, or stereoscopic matching, and so forth. Analogous pyramid schemes have been developed for computer vision. (See Adelson, E. H., Simoncelli, E., & Hingorani, R., 1987; Simoncelli and Olshausen, 2001). The second explanation is in terms of economical encoding which we pick up on briefly below.

It is interesting that many of early attempts to understand visual cortical receptive fields in terms of filters localized in space and spatial frequency were forerunners of modern wavelet theory (Simoncelli and Olshausen, 2001).

Edge Detection by Neural Networks

Because of the orientation selectivity of cortical cells, they have been sometimes interpreted as edge detectors. It is easy to see how a sine-phase Gabor function filter (1 cycle wide) would respond vigorously to an edge oriented with its receptive field. This type of receptive field behaves as a 2D smoothing operator followed by a first order directional derivative. There are various ways of "reverse engineering" the visual system to design edge detectors. Here is one that combines the outputs of two model simple cells.

Optional: Morrone and Burr edge detector: combining sine and cosine phase filters

What kinds of computational function might complex cells serve?

- Define the filters

```math
\text{cosinefilter}[x_,\sigma_,f_] := \exp[-(x/\sigma)^2] \cos[2 \pi f x]
\text{sinefilter}[x_,\sigma_,f_] := \exp[-(x/\sigma)^2] \sin[2 \pi f x]
```
- Define the input stimulus: an ideal edge

\[
\text{Plot}\left[\text{Sign}\left[x\right],\{x,-1,1\},\text{Axes}\to\text{None}\right]
\]

- Calculate the response of a bank of cosine filters to the edge

\[
cr[x_] := \text{NIntegrate}\{\text{cosinefilter}\left[(x - x1), .5, 1\right] \text{Sign}\left[x1\right],\{x1,-1.5,1.5\}\}
\]

\[
\text{Plot}\left[cr[x],\{x,-1,1\},\text{PlotPoints}\to10\right];
\]

- Calculate the response of a bank of sine filters to the edge

\[
sr[x_] := \text{NIntegrate}\{\text{sinefilter}\left[(x - x1), .5, 1\right] \text{Sign}\left[x1\right],\{x1,-1.5,1.5\}\}
\]
Plot[ sr[x], {x, -1, 1}, PlotPoints -> 10];

-1
-0.5
0.5
1
-0.1
-0.2
-0.3
-0.4

-1
-0.5
0.5
1
0.025
0.05
0.075
0.1
0.125
0.15

■ Combine the outputs from the two banks by squaring and adding:

Plot[ ch[x]^2 + sr[x]^2, {x, -1, 1}, PlotPoints -> 10]

The peak of the squared sum is at the location of the edge.

Note the basic structure of the neural network for edge detection:

1. A linear operation using weights from a cosine function, followed by a squaring operation (how does this squaring relate to the sigmoidal non-linearities we have been using?)

2. A linear operation using weights from a sine function, followed by a squaring operation

3. A linear sum of the outputs

Edges in the real world of natural images don't necessarily correspond to a sharp ideal edge, but can be blurry and noisy. Nevertheless they are perceived to be sharp.

Morrone and Burr went on to show that one could do the same operation with different sizes of filters (i.e. different values of sigma), and each time the peak of the above operation for an ideal edge occurs at the edge transition. But even for blurry edges, the larger scale filters will still find a point in the transition region. Thus by adding up a whole set of neural outputs over a range of scales, one could detect an edge. Another way of viewing this network is one that detects phase coherence. Fourier theory shows that a step function can be built up of sine-waves of various frequencies whose zero crossings all line up with (say positive slope) at the edge transition.

Although one can build edge detectors from oriented filters, simple cells cannot uniquely signal the presence of an edge for several reasons. One is that their response is a function of many different parameters. A low contrast bar at an optimal orientation will produce the same response as a bar of higher contrast at a non-optimal orientation. There is a similar trade-off with other parameters such as spatial frequency and temporal frequency. In order to make explicit the location of an edge from the responses of a population of cells, one would have to compute something like the "center-of-mass" over the population, where response rate takes the place of mass. Another problem is that edge detection has to take into
account a range of spatial scales. The cortical basis set does encompass a range of spatial scales, and in fact may be "self-similar" across these scales. See Koenderink (1990) for a recent theoretical discussion of "ideal" receptive field properties from the point of view of basis elements.

**Motion Detection by Neural Networks - detecting orientation in space-time**

Networks analogous to the Morrone-Burr detector have been proposed for detecting motion (Adelson and Bergen, 1985; Heeger et al, 1996; Simoncelli and Heeger, 1997).

**Economical representations by neurons in primary cortex**

We might expect something like Fourier analysis of the image to result in efficient coding because of the close relationship between Fourier rotations and Principal Components Analysis (e.g. Appendix A, Andrews, 1983). Fourier coefficients for natural images tend to be uncorrelated. Some work has been completed toward a functional explanation for the orientation and spatial frequency tuning properties of cortical receptive fields based on the statistics of natural images (Field, 1994), but the story is far from complete. Barlow has argued that a decorrelated representation of sensory information is important for efficient learning (Barlow, 1990).

There has been rapid progress in the relationship between self-organizing models of visual cortex, and efficient coding of image information. For more on this, see: Linsker, R. (1990), Barlow, H. B., & Foldiak, P. (1989), Olshausen and Field (1996), Simoncelli and Olshausn (2001). Linsker's computational studies show, for example, that orientation tuning, and band-pass properties of simple cells can emerge as a consequence of maximum information transfer (in terms of variance) given the constraint that the inputs are already band-pass, and the receptive field connectivity is a priori limited.

Olshausen and Field (1996) show that simple cell receptive field weights can emerge as a consequence of sparse coding. Imagine you don't know what the receptive field weights should be, and would like to write an algorithm to discover what the should be given simple principles. Olshausen and Field showed that if one enforces two constraints: 1) the image should be well-approximated by a weighted some of basis vectors; 2) when representing natural images, the (squared) coefficients (i.e. the projections of the image vector onto the basis vectors) should add up to a small number, then the receptive field weight structures emerge as a consequence.

We'll take a closer look at the topic of neural networks for efficient encoding in the next lecture (Lecture 14).

**Extra-striate cortical visual areas**

In lecture 1, we saw that cells in the visual cortex send their visual information to an incredibly complex, and yet structured collection of extra-striate areas. Any hypothesized function of striate cortex must eventually take into account what the information is to be used for. Two primary functions: object perception and recognition--"within object" processing, and spatial processing (between object, and view-object relations).
Spatial, action pathway

V1, MT, MST, LIP, ...

Viewer-centered computations

"Where" (vs. "What")

("where" or "how" or "now")

Spatial computations, such as coordinate transformations for action

Object perception, recognition pathway ("what")

object perception, recognition pathway

V1, V2, V4, Posterior IT, Anterior IT, ... - generic feedforward neural network models PLUS feedback

Invariances required for recognition:

photometric: illumination level, direction, shadows

geometrical: translation, size, orientation in depth

category-related: levels of abstraction

Binding problem:

grandmother cells, distributed codes, sparse codes - "binding by synchrony"
Modeling large-scale neural systems & systems analysis

Much of the modeling of visual processing has been built on the tools that we've learned about. But there are many aspects of brain modeling that require additional tools and ways of thinking.

Modeling information processing functions:

Feedback
   Information processing roles of feedback
   *Dynamical behavior
Timing and sequences (e.g. speech, motor sequences)
Dynamical issues for real-time control, visuo-motor control
Large-scale architectures (e.g. Inter-area processing)
   *Handling uncertainty - Probabilistic models

Measuring and characterizing neural systems

Linear and non-linear systems analysis, statistical and stochastic processes analysis (time series),...
References


National Academy of Science. 93, 623–627.


1193-1216.
