

Neural Encoding II: Reverse Correlation and Visual Receptive Fields

Theoretical Neuroscience Ch 2 Notes (Dayan and Abbott)

Neha Peddinti

November 2021

Contents

1	Goal	2
2	Estimating firing rates	2
3	Early visual system anatomy & physiology	3
4	Reverse-correlation methods: simple cells	5
5	Static nonlinearities: complex cells	7
6	Receptive fields in the retina and LGN	8
7	Constructing V1 receptive fields	8
8	Summary	8

1 Goal

1. Use spike-triggered averages (STAs) of stimuli, along with reverse-correlation techniques, to estimate neuron firing rates caused by time-dependent stimuli.
2. Use different models to account for nonlinear effects of neuron firing rate changes, like firing thresholds.
3. STAs and reverse-correlation techniques are generally useful to figure out how visual neurons (retina, LGN, primary visual cortex V1) encode info, so most examples in the chapter are from those. Can't really capture how the nonlinear, more complex features of responses encode info.
4. Basic method of simulating how neurons are **encoded** (stimulus to neuron response):
Stimulus function (input) → Linear filter on stimulus → Static nonlinearity filter → Spike generator (use the resulting firing rate function to generate spikes according to that probability of spiking) → Response (the final spike train output)
5. None of these techniques says anything about how anything else besides the stimulus affects the neuron response (synaptic, cellular, and network properties of the neuron circuits), so the models are rather simple, but they're useful in some cases and for crude predictions.

2 Estimating firing rates

1. Summary

From Ch 1: Plotted neural response vs. stimulus (stimulus tuning curve) and used that as a model of firing rate as an instantaneous function of the stimulus.

Now: Take into account the stimulus a few hundred ms before t until tens of ms before t to model the neuron's activity at time t .

2. Reverse correlation: used to study linear response characteristics of single neurons; estimates the firing rate by weighting all the past stimulus values with some function D .
3. Nonlinear filter/nonlinear effects: output does not vary linearly with input (scaling the input up by some factor doesn't necessarily scale the output up by the same factor, etc.)
4. Linear filter D on 2 signals A and B : $D(A + \lambda B) = D(A) + \lambda D(B)$ should be true (so the filter can be applied with convolution):

$$r_{\text{est}}(t) = r_0 + \int_0^\infty d\tau D(\tau) s(t - \tau) \quad (1)$$

where r_0 represents any constant background firing, $s(t - \tau)$ is the stimulus value τ ms before t , and $D(\tau)$ is the weighting factor used to determine how much the stimulus value τ ms before t effects the firing rate.

Note: Volterra or Wiener expansions have more terms than this and account for the correlation between more than 2

times at once, and might be useful to capture some nonlinear effects, but this is good enough for now.

Assume that $s(t)$ has been defined so that its mean is 0 for further calculation purposes (as it can always be shifted accordingly so that this is the case).

5. Error of the above estimate

$$E = \frac{1}{T} \int_0^T dt (r_{\text{est}}(t) - r(t))^2 \quad (2)$$

Set derivative with respect to the function D to 0 to find D such that error E is minimized. This D is the "optimal kernel."

6. Optimal kernel derivation:

see **Appendix A for derivation until this point**

The function D that minimizes error is

$$\int_0^\infty d\tau' Q_{ss}(\tau - \tau') D(\tau') = Q_{rs}(-\tau) \quad (3)$$

where

$$Q_{ss}(\tau) = \int dt s(t) s(t + \tau)/T \quad (4)$$

is the stimulus autocorrelation function (used to see if info is encoded in the timings of spikes relative to other spikes), and

$$Q_{rs}(\tau) = \int dt r(t) s(t + \tau)/T \quad (5)$$

is the firing rate-stimulus correlation function.

Note: The method is called *reverse correlation* because of the $-\tau$ in the equation.

Example: The optimal kernel equation can be solved easily if the stimulus is white noise because then: $Q_{ss}(\tau) = \sigma_s^2 \delta(\tau)$, so:

$$\sigma_s^2 \int_0^\infty d\tau' \delta(\tau - \tau') D(\tau') = \sigma_s^2 D(\tau) = Q_{rs}(-\tau) \quad (6)$$

So, the kernel with the best linear estimate of the firing rate is:

$$D(\tau) = \frac{Q_{rs}(-\tau)}{\sigma_s^2} = \frac{\langle r \rangle C(\tau)}{\sigma_s^2}, \quad (7)$$

where $C(\tau)$ is the STA and $\langle r \rangle$ is the average firing rate of the neuron (the relation $Q_{rs}(-\tau) = \langle r \rangle C(\tau)$ is derived in Ch. 1).

This is the white-noise kernel, which measures the STA in response to a white-noise stimulus to determine the optimal kernel D with the equation above.

7. Most effective stimulus

Use the optimal kernel to find the stimulus that is predicted to result in the max firing rate.

Problem: stimulus can be scaled up indefinitely for the integral with the linear filter to become unbounded. → There's no theoretical "max."

Solution: constrain stimulus to have a constant **stimulus energy** for all possible stimulus functions (time integral of the square of the stimulus is fixed). Stimulus with max firing rate response with this constraint is the one that the neuron probably selects for.

see appendix B for derivation of most effective stimulus

End result is that the most effective stimulus is proportional to the optimal kernel D .

8. Intuitive interpretation:

Linear filter (Equation 1) is proportional to the convolution of the stimulus with the most effective stimulus (the one the neuron encodes for), so the firing rate estimate is proportional to the overlap between the two.

9. Shortcomings of model: H1 neuron of fly: responds to moving images.

H1 neuron data for the response to a moving visual image was compared with the optimal kernel firing rate estimate. Graphs did not align when the firing rate had a higher frequency, possibly due to nonlinear effects. → Linear filter can't account for these.

Problems with linear filter include:

- (a) Negative firing rate is possible.
- (b) Predicted firing rate increases without bound as the magnitude of the stimulus increases (no firing threshold or saturation)

10. Solution (static nonlinearities): static because it's a function of the linear filter value, so it's evaluated instantaneously.

Linear filter L :

$$L(t) = \int_0^\infty d\tau D(\tau)s(t - \tau) \quad (8)$$

Transform linear prediction L to account for above problems, and add a constant background firing rate:

$$r_{\text{est}}(t) = r_0 + F(L(t)) \quad (9)$$

where F is determined from data (find curve of best fit by plotting the observed r vs L for various times and with various stimuli).

- (a) Half-wave rectification ($[-]_+$): sets all negative values to 0.
- (b) Threshold function: L must hit threshold value L_0 before firing.

$$F(L) = G[L - L_0]_+$$

- (c) Saturation (function asymptotically approaches some upper bound to prevent firing rate prediction from increasing too much):

Sigmoidal function, where r_{max} is the max value, $L_{1/2}$ is the L at which F achieves half of this value, and g_1 determines the steepness of the function.

$$F(L) = \frac{r_{\text{max}}}{1 + \exp(g_1(L_{1/2} - L))}$$

Another option:

$$F(L) = r_{\text{max}}[\tanh(g_2(L - L_0))]$$

11. Problems with static nonlinearities

- (a) Doesn't allow for dependence of F on autocorrelations of the stimulus or other terms of the Volterra expansion.
- (b) Linear kernel D derived before is no longer optimal since it was optimized to reduce the error of the linear estimate.

see Bussgang's theorem, appendix C to show that D will still be reasonable if the white-noise stimulus used is Gaussian.

3 Early visual system anatomy & physiology

1. Converting a light stimulus to a spike train:

- (a) Phototransduction in photoreceptors (rods and cones) at the back of the retina triggers electrical signal after receiving light.
- (b) Photoreceptors are hyperpolarized by light (membrane potential codes for light intensity, which works for small distances).
- (c) Axons of ganglion cells (final stage of the retinal circuit) form optic nerve → neurons cross paths at optic chiasm → LGN → Area V1
- (d) LGN and V1 neurons individually respond to light stimuli in their respective receptive fields (can be indirectly influenced by light outside of their receptive fields): some cells sum contributions from different parts of the receptive field linearly, and some do not (nonlinear).

2. Receptive fields:

- (a) Retinal ganglion cells/LGN neurons: circular spots of light surrounded by darkness or circles of darkness surrounded by light.
- (b) V1 neurons: elongated light/dark bars or boundaries between light/dark regions (like gratings)

3. Retinotopic map: transformation of real (visual) world coordinates to cortical surface locations (because apparently neurons that encode nearby receptive fields tend to be physically near each other in the cortex).

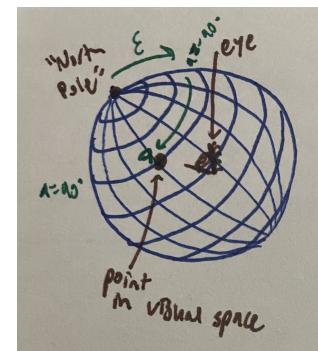


Figure above: Eccentricity ϵ is like latitude and azimuth a is longitude on a sphere where the "North Pole" is the point that focuses into the center of your retina (visual world coordinates). Depth (radius of sphere) doesn't matter bc a single eye can only determine coords ϵ and a .

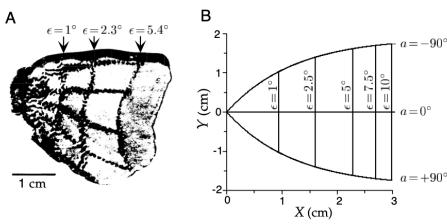


Figure above: Transformation to Cartesian (cortical) coords. Fig. A is a scan of the posterior primary visual cortex (left side) of a monkey shown a set of concentric circles similar to grid lines of a polar coordinate system (centered at "North Pole" of retina map). Neural activity locations matched grid lines: vertical lines are the concentric circles, horizontal-ish lines are the radial lines. Fig. B shows the approximate mapping.

Cortical magnification factor $M(\epsilon)$: encodes relationship between angular displacement in visual world and cortical displacement. Just a function of ϵ , which seems intuitive based on the above figure.

$$\Delta X = M(\epsilon) \Delta \epsilon \quad (10)$$

$$\Delta Y = M(\epsilon) \theta \quad (11)$$

where θ is the angular distance between two points with the same ϵ caused by a small change in a .

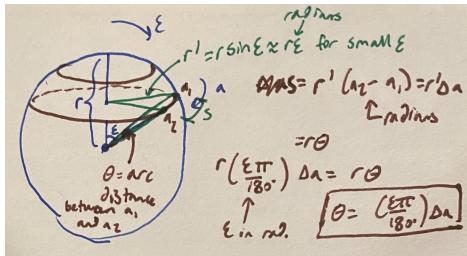
$$\frac{dX}{d\epsilon} = M(\epsilon)$$

Curve of best fit from points suggests:

$$M(\epsilon) = \frac{\lambda}{\epsilon_0 + \epsilon}$$

with $\lambda \approx 12$ mm and $\epsilon_0 \approx 1^\circ$. So integrating (with $X = 0$ representing $\epsilon = 0$ gives:

$$X = \lambda \ln(1 + \epsilon/\epsilon_0)$$



From above diagram:

$$\theta = \left(\frac{\epsilon \pi}{180^\circ} \right) \Delta a$$

which means that

$$\Delta Y = M(\epsilon) \theta = M(\epsilon) \left(\frac{\epsilon \pi}{180^\circ} \right) \Delta a \quad (12)$$

$$\frac{dY}{da} = -\frac{\epsilon \pi}{180^\circ} M(\epsilon) \quad (13)$$

$$Y = -\frac{\lambda \epsilon a \pi}{(\epsilon_0 + \epsilon) 180^\circ} \quad (14)$$

where the minus sign appears because the visual field is inverted on the cortex. These give you the map in Fig. B from earlier.

4. Cortical map as a complex logarithmic map: For large ϵ :

$$X \approx \lambda \ln \left(\frac{\epsilon}{\epsilon_0} \right)$$

$$Y \approx -\frac{\lambda \pi a}{180^\circ}$$

so we can define

$$z = \left(\frac{\epsilon}{\epsilon_0} \right) \exp(-i\pi a/180^\circ) \quad (15)$$

$$Z = \lambda \ln(z) = X + iY \quad (16)$$

Note that if you scale an image radially by factor γ :

$$\epsilon \rightarrow \gamma \epsilon$$

$$X \rightarrow \lambda \ln \left(\frac{\gamma \epsilon}{\epsilon_0} \right) \rightarrow X_\lambda \ln(\gamma)$$

so scaling radially affects the X coord of the responding neuron in the cortex. a is unchanged, so Y is unchanged.

5. Visual stimulus functions

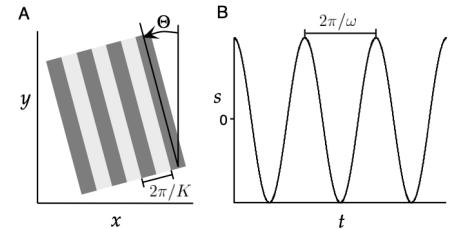
(a) Generally only measures contrast (luminance minus background luminance, divided by background luminance to make it dimensionless)

(b) Counterphase sinusoidal grating:

$$s(x, y, t) = A \cos(Kx \cos \Theta + Ky \sin \Theta - \Phi) \cos(\omega t)$$

where K and w are spatial and temporal (angular) frequencies of the grating, and A is contrast amplitude (max difference between light and dark areas). Θ is generally measured with respect to the y axis so that a stimulus with $\Theta = 0$ varies in the x direction.

x and y are in degrees so K is in rad/deg.



(c) White-noise image would satisfy

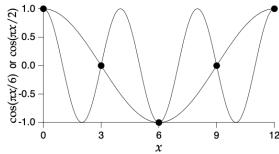
$$\frac{1}{T} \int_0^T dt s(x, y, t) s(x', y', t+\tau) = \sigma_s^2 \delta(\tau) \delta(x-x') \delta(y-y')$$

using techniques from Ch 1, including the zero correlation with time definition of white noise.

6. Nyquist frequency: max spatial (angular) frequency that can be fully detected (resolved) by the visual system. Densely-packed photoreceptors in an array sample points at locations $m\Delta x$ for $m = 1, 2, \dots$, and the frequency defining the resolution of this array is

$$K_{\text{nyq}} = \frac{1}{2} \cdot \frac{2\pi}{\Delta x} = \frac{\pi}{\Delta x} \quad (17)$$

because each half wavelength of the resolved visual stimulus (the thing you're trying to detect) will need 3 detected points (end points of a full wavelength of the sampling frequency wave) to be fully resolved.



4 Reverse-correlation methods: simple cells

1. Some functions that just add spatial dependence to the functions for time-dependent stimuli derived earlier:

(a) STA

$$C(x, y, \tau) = \frac{1}{\langle r \rangle} \left\langle \sum_{i=1}^n s(x, y, t_i - \tau) \right\rangle$$

(b) Correlation between firing rate at time t and stimulus at time $t + \tau$

$$Q_{rs}(x, y, \tau) = \frac{1}{T} \int_0^T dt r(t) s(x, y, t + \tau)$$

(c) Relating the two:

$$C(x, y, \tau) = \frac{Q_{rs}(x, y, -\tau)}{\langle r \rangle}$$

2. Space-time receptive field (RF): Contributions to linear estimate of firing rate at time T come from weighting different spatial and temporal locations of the stimulus prior to t . Simple linear estimate is

$$L(t) = \int_0^\infty d\tau \int dx dy D(x, y, \tau) s(x, y, t - \tau)$$

where $D(x, y, \tau)$ is the kernel, which determines the magnitude and sign of the influence that the visual stimulus at $(x, y, t - \tau)$ affects the firing rate at t .

3. Optimal kernel: minimizes error, which is defined to be the magnitude of the difference in the function D and the observed firing rate function.

With Ch 1 techniques, optimal kernel is the STA for a white-noise stimulus with variance σ_s^2 :

$$D(x, y, \tau) = \frac{Q_{rs}(x, y, -\tau)}{\sigma_s^2} = \frac{\langle r \rangle C(x, y, \tau)}{\sigma_s^2}$$

4. Separable/nonseparable receptive fields:

If the preferred stimulus of the cell maintains its locations and shapes over time (with just magnitudes and signs of stimulus in different spatial regions changing over time), this indicates that it can be approximated with a separable space-time receptive field (see Nonseparable RFs section for explanation):

$$D(x, y, \tau) = D_s(x, y) D_t(\tau)$$

5. Spatial receptive fields

- (a) ON/OFF regions: areas where D_s is positive (light leads to a positive contribution to the firing rate) or negative (darkness leads to a positive contribution to the firing rate).

- (b) Gabor function to model that - multivariate normal dist weighted by cos function:

$$D_x(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2}\right) \cos(kx - \phi)$$

where k is the preferred spatial frequency of the grating (visual stimulus) and ϕ is the preferred spatial phase.

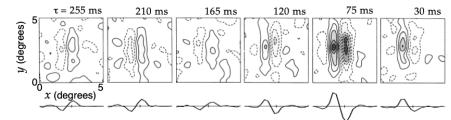
Change preferred orientation angle θ by making the substitutions $x \rightarrow x \cos \theta + y \sin \theta$ and $y \rightarrow y \cos \theta - x \sin \theta$.

- (c) Bandwidth: width of spatial frequency tuning curve measured in octaves (so small bandwidth is more selective).

$$b = \log_2 \left(\frac{k\sigma_x + \sqrt{2 \ln 2}}{k\sigma_x - \sqrt{2 \ln 2}} \right)$$

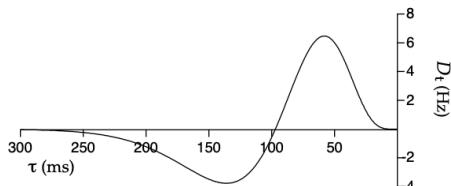
6. Temporal receptive fields

- (a) Example: preferred stimulus might be one that starts with an appropriately aligned light-dark pattern and then reverses to a dark-light pattern over time (diagram)



- (b) Two-phase response example (Adelson and Bergen - 1985): possible temporal structure of the RF

$$D_t(\tau) = \alpha \exp(\alpha\tau) \left(\frac{(\alpha\tau)^5}{5!} - \frac{(\alpha\tau)^7}{7!} \right)$$



7. Putting everything together: response of a simple cell to a counterphase grating

- (a) Assuming approximately separable receptive fields, the linear estimate of response can just be:

$$L(t) = L_s L_t(t).$$

For a stimulus that's just a moving counterphase grating (so a grating that changes intensity in a given location with time) with spatial phase Φ , orientation (Θ), spatial frequency K , and temporal frequency ω :

$$L_s = \int dx dy D_x(x, y) A \cos(Kx \cos \Theta + Ky \sin \Theta - \Phi)$$

and

$$L_t(t) = \int_0^\infty d\tau D_t(\tau) \cos(\omega(t - \tau))$$

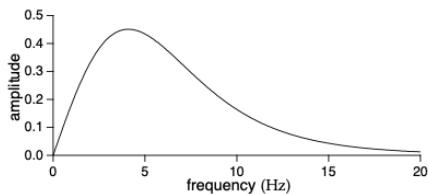
- (b) Ex: L_s for $\Phi = 0$, where Φ is the preferred spatial phase

$$L_s = A \exp\left(-\frac{\sigma^2(k^2 + K^2)}{2}\right) \cosh(\sigma^2 k K \cos \Theta)$$

- (c) Ex: L_s when $\Theta = 0$ of preferred counterphase grating

$$L_s = \frac{A}{2} \exp\left(-\frac{\sigma^2(k - K)^2}{2}\right) \cos(\phi - \Phi)$$

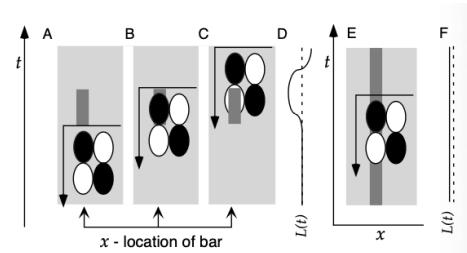
- (d) Ex: frequency response of model simple cell from the L_t (firing rate) model based on the D_t in temporal receptive fields section; firing rate follows sinusoidal pattern, but the amplitude of the frequency response (the max firing rate) changes depending on how close the counterphase grating temporal frequency is to the preferred frequency. Peak value of 4 Hz is typical for V1 neurons.



8. Space-time receptive fields

- (a) Drop y-spatial dependence to graph space-time diagram of both the preferred stimulus and the actual stimulus.
 (b) Static bar generates no response because of overlap on both ON and OFF regions (E and F of diagram).

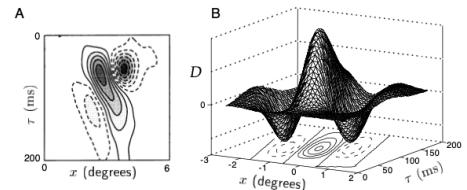
(c) Moving grating is an effective stimulus (A-D of diagram)). Note that in the figures, the RF itself also has a time component, but since the space-time diagram extends much past the max τ of the RF, the RF is plotted as moving in the direction of increasing time as well, and the overall alignment of both space and time with the grating bars determines $L(t)$.



- (d) Separable space-time RF is not sensitive to direction of motion because it is only sensitive to light-dark oscillations, which happen regardless of the grating's direction of motion.

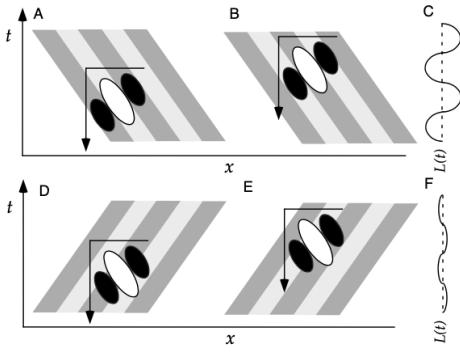
9. Nonseparable receptive fields

- (a) Required to detect direction of motion.
 (b) Ex: Figure shows 3-lobed OFF-ON-OFF spatial RF (from cat primary visual cortex), and you can see that the regions move to the left with time, so its optimal stimulus is a grating moving to the left.



Nonseparable because the spatial location of the optimal stimulus depends on time (if only the amplitude of the optimal stimulus changes with time, the temporal response can be encoded by D_t separately for separable space-time RFs since D_s would still encode for the same stimulus max/min locations; but that's not the case here since the optimal stimulus max/min locations also change with time, so the phase of D_s would have to change with time and would no longer be time-independent).

- (c) To clarify how the direction of motion impacts the firing rate, notice the difference of incorporating 3 regions of variable maxima/minima locations:



- (d) Describing nonseparable RF structure by transforming the space and time coordinates:

$$D(x, y, \tau) = D_s(x', y)D_t(\tau')$$

$$x' = x \cos \psi - c\tau \sin \psi$$

$$\tau' = \tau \cos \psi + \frac{x}{c} \sin \psi$$

Transformation with x' and τ' and justification: it essentially just rotates a separable space-time RF inn the space-time diagram (in the xt plane), where ψ is the $x - \tau$ rotation angle in the space-time diagram.

10. Static nonlinearities: simple cells

- (a) Neural responses saturate (stop getting really large) when the image contrast increases, and response can't be negative, so Heeger (1992) proposed this static nonlinearity to apply to the linear firing rate estimate L :

$$F(L) = \frac{G[L]_+^2}{A_{1/2}^2 + G[L]_+^2}$$

where G is set to fit the magnitude of the measured firing rates. F is the new firing rate estimate, but note that D will no longer be the kernel that minimizes error once you apply this new function to it.

5 Static nonlinearities: complex cells

1. Complex cell: You can't first approximate the response with a linear function and then apply a static nonlinearity to correct certain features; instead the cell simply can't be divided into regions that sum linearly to create a response, so linear functions are incompatible with the cell's features. Also means that STA would be hard to interpret.

2. Properties

- (a) Spatial-phase invariance: cell responds to light/dark bars regardless of position in RF (selects for an image regardless of exact spatial position within RF), possibly an early stage of position-invariant object recognition.
- (b) Constant response to moving grating, not oscillatory response (which was the case for simple cells).

- (c) Frequency doubling: complex cell responses to grating oscillating with frequency ω has a constant component and an oscillatory component with frequency 2ω (see next section for explanation).

3. Deriving the spatial-phase-invariant response of a complex cell (to get a function that does not depend on the Θ or Φ of the stimulus grating).

For L_1 and L_2 of preferred spatial phases ϕ and $\phi - \pi/2$, you can derive:

$$L_1 = AB(\omega, K) \cos(\phi - \Phi) \cos(\omega t - \delta)$$

$$L_2 = AB(\omega, K) \sin(\phi - \Phi) \cos(\omega t - \delta)$$

$$L_1^2 + L_2^2 = A^2 B^2(\omega, K) \cos^2(\omega t - \delta)$$

$$r(t) = r_0 + G(L_1^2 + L_2^2)$$

where $B(\omega, K)$ is some factor that you can find the explicit function for using the definition of L explained earlier, $r(t)$ is the final response spatial-phase-invariant response estimate, r_0 is some constant background firing rate, and G is some constant.

Note that the frequency-doubling effect is included in this model because the $\cos^2(\omega t - \delta)$ factor in $L_1^2 + L_2^2$ has frequency 2ω .

4. Fig. 2.24: Compare simple and complex cell responses.

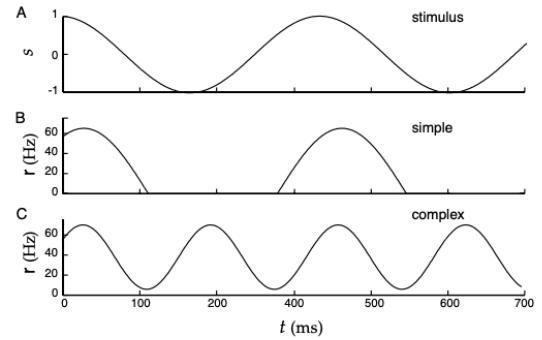


Figure 2.24 Temporal responses of model simple and complex cells to a counter-phase grating. (A) The stimulus $s(x, y, t)$ at a given point (x, y) plotted as a function of time. (B) The rectified linear response estimate of a model simple cell to this grating with a temporal kernel given by equation 2.29 with $\alpha = 1/(15 \text{ ms})$. (C) The frequency-doubled response of a model complex cell with the same temporal kernel but with the estimated rate given by a squaring operation rather than rectification. The background firing rate is $r_0 = 5 \text{ Hz}$. Note the temporal phase shift of both B and C relative to A.

5. Energy model: Complex cell model above is similar to simple harmonic oscillator equation because the pair of linear filters L_1 and L_2 are separated by a spatial phase of $\pi/2$ (quadrature pair), so it's also called the energy model.

6. Note: You can write the complex cell response as the sum of squares of four rectified simple cell responses (which should be fairly intuitive from the graph).

$$r(t) = r_0 + G([L_1]_+^2 + [L_2]_+^2 + [L_3]_+^2 + [L_4]_+^2)$$

where the simple cells have 4 different preferred spatial phases ϕ , $\phi + \pi/2$, $\phi + \pi$, and $\phi + 3\pi/2$. \rightarrow indication about how complex cells take into account all orientations Φ equally, but also note that this is NOT an indication that the complex cell gets input from 4 simple cells.

6 Receptive fields in the retina and LGN

1. Prior discussion was only for V1 neurons.

2. Some retinal ganglion cells (X cells in cats, P cells in monkey retina/LGN) can be modelled with reverse-correlation methods.

Receptive fields have circular central ON region surrounded by disk OFF region (instead of ON/OFF bar pattern of V1 neurons).

Possible spatial RF (with center of RF at $x = y = 0$):

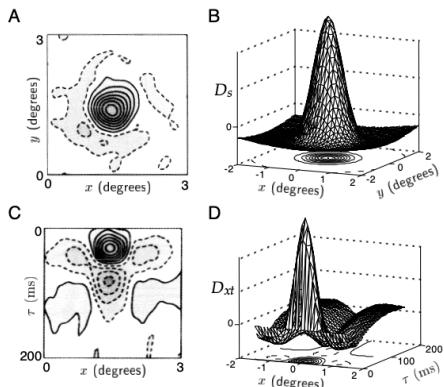
$$D_s(x, y) = \pm \left(\frac{1}{2\pi\sigma_{cen}^2} \exp\left(-\frac{x^2 + y^2}{2\sigma_{cen}^2}\right) - \frac{B}{2\pi\sigma_{sur}^2} \left(-\frac{x^2 + y^2}{2\sigma_{sur}^2}\right) \right)$$

where σ_{cen} controls size of central region, σ_{sur} controls size of disk region, B controls balance of contributions from the two regions.

Nonseparable RF: This is because reversal of surrounding ring region (for the preferred stimulus) is general slower than the reversal of the central region, so the temporal kernel D_t depends on which location you're observing.

$$D(x, y, \tau) = \pm \left(\frac{D_t^{cen}(\tau)}{2\pi\sigma_{cen}^2} \exp\left(-\frac{x^2 + y^2}{2\sigma_{cen}^2}\right) - \frac{BD_t^{sur}(\tau)}{2\pi\sigma_{sur}^2} \left(-\frac{x^2 + y^2}{2\sigma_{sur}^2}\right) \right)$$

3. Possible LGN neuron receptive field:



7 Constructing V1 receptive fields

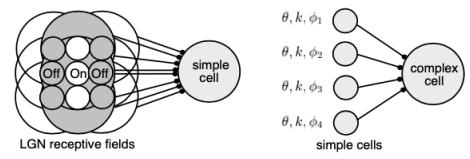
1. Hubel-Wiesel (1962) single cell model: LGN neurons code ON/OFF for different locations, and those are summed together to form oriented RFs in cortical (V1) neurons

2. Hubel-Wiesel complex cell model: single cells with similar θ and k preferences but different ϕ preferences are inputs to a

single complex cell.

Note that V1 neurons are likely to have other inputs besides simple cells, but that isn't accounted for in this model.

3. Diagram of the models:



Summary

1. The RF (receptive field) is the kernel used in the linear filter L , which represents the first approximation of the firing rate when using reverse-correlation methods. This model only works for simple cells, which sum the weighted contributions from different positions of the RF.
2. V1 (cortical) simple cells can generally be modeled by space-time RFs that have vertical bars as ON/OFF regions. These models are confirmed by experiment.
 - (a) Separable RFs can detect the motion of counterphase gradients, but not direction.
 - (b) Nonseparable RFs can detect direction as well, in general due to their RF having 3 regions that move left to right (or right to left) on a space-time diagram.
3. V1 complex cell responses are not dependent on stimuli position or orientation within the RF, so they cannot be modeled by a linear filter because contributions from different positions are not weighted and summed. It can be shown that a complex cell response can be written as a function of the sum of squares of the linear responses of simple cells that prefer different spatial orientations.
4. Some retinal ganglion and LGN neurons have an RF with a central ON region and a surrounding disk OFF region (or vice versa), but similar methods can be applied.
5. The Hubel-Wiesel models of RFs suggest that LGN neurons that classify each location as ON or OFF are the inputs to simple cells, which might be how an oriented RF is implemented in neural circuitry.

Note: All figure credits go to *Theoretical Neuroscience* (Dayan and Abbott, Ch. 2), except for my hand-drawn figures.