Lateral organization & computation

Population encoding & decoding

review

lateral organization

Retinotopic maps

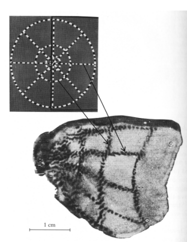
- Log-polar model
  - see: smallRetinaCortexMap.nb

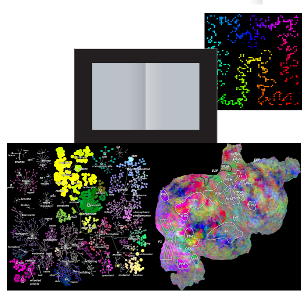
Other maps? Grouping what?

- http://gallantlab.org/publications/huth-et-al-2012.html
- http://gallantlab.org/semanticmovies/

Efficient representations that reduce or exploit redundancy

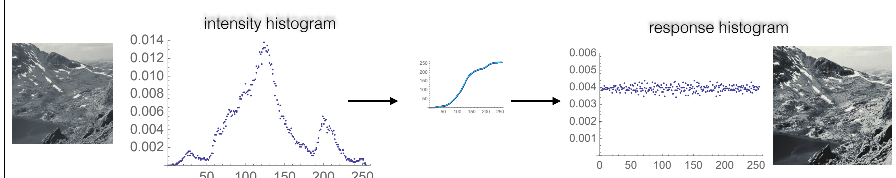
- sparse coding theories. “dictionary” methods





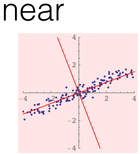
Efficient representations that reduce or exploit redundancy

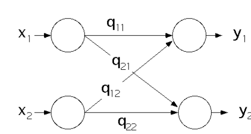
1rst order



2nd order, linear

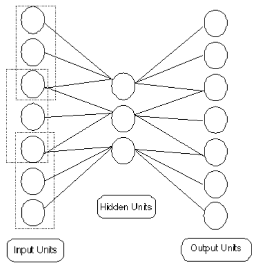
PCA





but needs modified Oja rule to capture all components: $\Delta q_{ij} = \alpha\left(x_j y_i - y_i \sum_{k=1}^{i} q_{kj} y_k\right)$

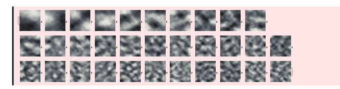
"autoencoder networks"



L ~ L'

Efficient representations that reduce or exploit redundancy

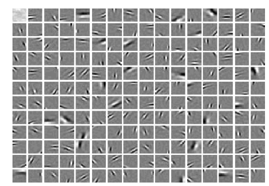
2nd order



$$I(x,y) = \sum_{i=1}^{n} A_i(x,y) s_i$$

PCA is a linear transform that decorrelates the coefficients:

$E(s_i s_j) = E(s_i) E(s_j)$

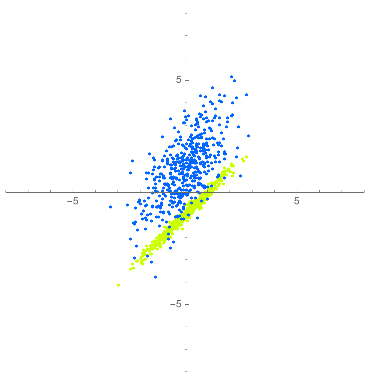


$$[L(x,y) - \sum_i s_i A_i(x,y)]^2 + \sum_i B(s_i)$$

ICA finds a linear decomposition such that: $p(s_i, s_j) = p(s_i) p(s_j)$

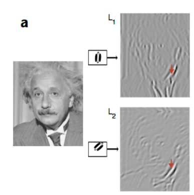
Hyvärinen, A. (2010). Statistical Models of Natural Images and Cortical Visual Representation. *Topics in Cognitive Science*, *2*(2), 251–264. doi:10.1111/j.1756-8765.2009.01057.x

PCA vs. Linear Discriminant Analysis



from lecture 18

Higher-order structure?



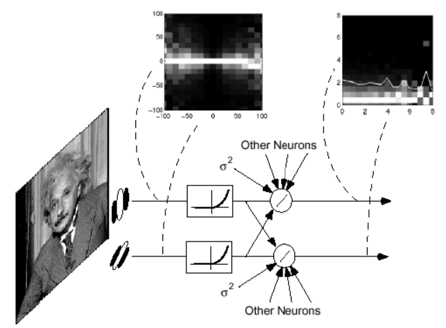


Figure 1: Illustration of image statistics as seen through two neighboring receptive fields. Left image: Joint conditional histogram of two linear coefficients. Pixel intensity corresponds to frequency of occurrence of a given pair of values, except that each column has been independently rescaled to fill the full intensity range. Right image: Joint histogram of divisively normalized coefficients (see text).

responses of linear model neurons with receptive fields that are close in space, preferred orientation or spatial frequency are not statistically independent

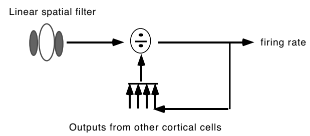
Schwartz, O., & Simoncelli, E. P. (2001). Natural signal statistics and sensory gain control. Nature Neuroscience, 4(8), 819–825.

Higher-order structure?

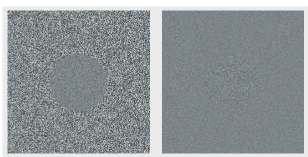
Accounts for neurophysiological responses of neurons in V1.

Schwartz, O., & Simoncelli, E. P. (2001). Natural signal statistics and sensory gain control. Nature Neuroscience, 4(8), 819–825.

divisive normalization



$$R_i = \sigma\left(\sum_{j=1}^{n} w_{ij} L_j\right) \bigg/ \sum_{k \in N_i} R_k^2$$



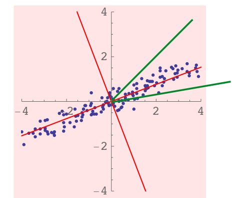
From Heeger

The middle disks have the same physical luminance variance, but the one on the right appears more "contrasty", i.e. to have higher variance.

This may be a behavioral consequence of an underlying non-linearity in the spatial filtering properties of V1 neurons involving "divisive normalization" derived from measures of the activity of other nearby neurons.

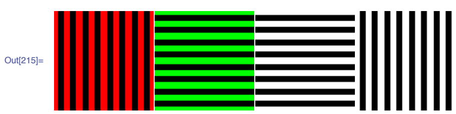
More on decorrelation:

non-orthogonal decorrelation

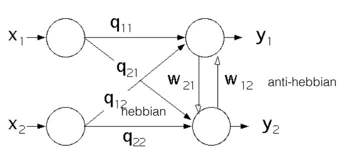


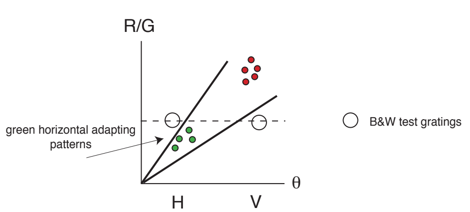
orthogonal orthogonal

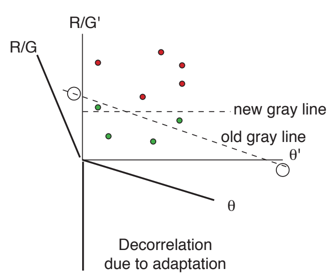
Contingent Adaptation



McCollough, C. (1965, 3 September 1965). Color Adaptation of Edge-Detectors in the Human Visual System. Science, 149, 1115-1116.







Decorrelation due to adaptation

Barlow, H. B., & Foldiak, P. (1989). Adaptation and decorrelation in the cortex. In C. Miall, R. M. Durban, & G. J. Mitchison (Ed.), The Computing Neuron Addison-Wesley.

ContingentAdaptation.nb

Lateral organization & neural codes

How do neural populations represent information?

Working assumptions:

Lateral organization involves a population of neurons representing features at the same level of abstraction

Receptive fields organized along a topographically mapped dimension with overlapping selectivities

Decoding — inferring world property from spikes— requires extracting information from the population

Mathematica notebook

Lect_24b_VisualRepCode.nb

Neural Implementations of Bayesian Inference

Lecture notes adapted from Alexandre Pouget
http://cms.unige.ch/neurosciences/recherche/

Zemel, R. S., Dayan, P., & Pouget, A. (1998). Probabilistic interpretation of population codes Neural Computation, 10(2), 403–430.

Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. Nature Neuroscience, 9(11), 1432–1438. doi:10.1038/nn1790

Probabilistic brains: knowns and unknowns (2013.) Pouget, A., Beck, J., Ma, W.J., Latham, P. Nature Neuroscience 16:1170-1178.

Perceptual encoding:
learning to represent world properties in terms of firing patterns

Perceptual decoding:
interpretation of encoded pattern by subsequent neural processes

Poisson noise

Imagine the following process: we bin time into small intervals, δt. Then, for each interval, we toss a coin with probability, P(head) =p. If we get a head, we record a spike. This is the Bernoulli process of PS#1.

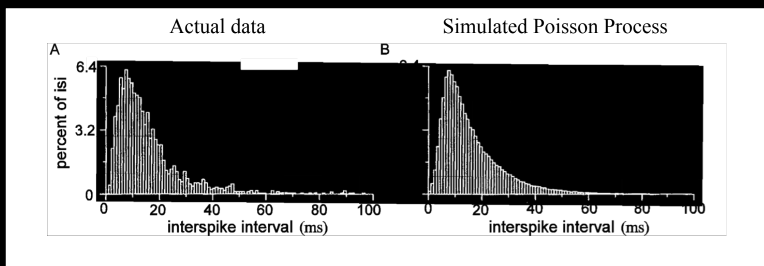
For small p, the number of spikes per second follows a Poisson distribution with mean $p/\delta t$ spikes/second (e.g., p=0.01, δt=1ms, mean=10 spikes/sec).

Properties of a Poisson process

- The variance should be equal to the mean
- A Poisson process does not care about the past, i.e., at a given time step, the outcome of the coin toss is independent of the past (“renewal process”).
- As a result, the inter-event intervals follow an exponential distribution (Caution: this is not a good marker of a Poisson process)

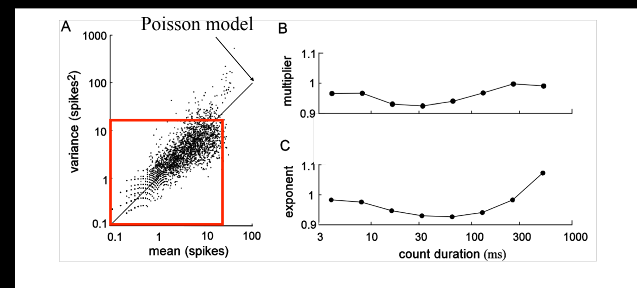
Poisson process and spiking

The inter spike interval (ISI) distribution is close to an exponential except for short intervals (refractory period) and for bursting neurons



Poisson process and spiking

The variance in the spike count is proportional to the mean but the the constant of proportionality can be higher than 1 and the variance can be an polynomial function of the mean. Log σ^2 = β Log a +log α



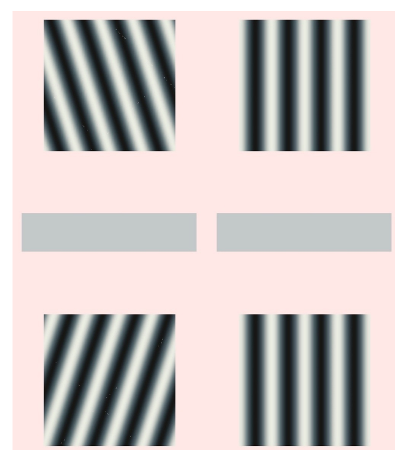
Is Poisson variability really noise?

Where could it come from?

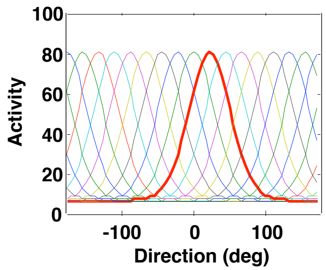
Neurons embedded in a recurrent network with sparse connectivity tend to fire with statistics close to Poisson (Van Vreeswick and Sompolinski, Brunel, Banerjee)

Could Poisson variability be useful for probabilistic computations? I.e. where knowledge of uncertainty is represented and used?

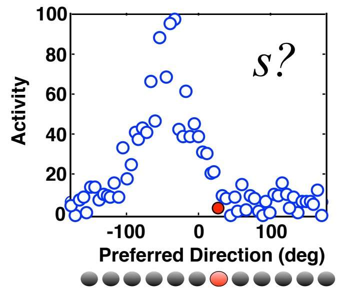
Poisson-like representations can be used for Bayesian integration of information



Population Code



Tuning Curves



Pattern of activity **(r)**

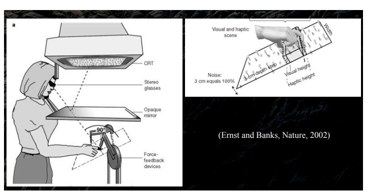
The decoding problem

Given a stimulus with unknown orientation s, what can one say about s given a vector r representing the pattern of neural activity?

Estimation theory: come up with a single value estimate from r

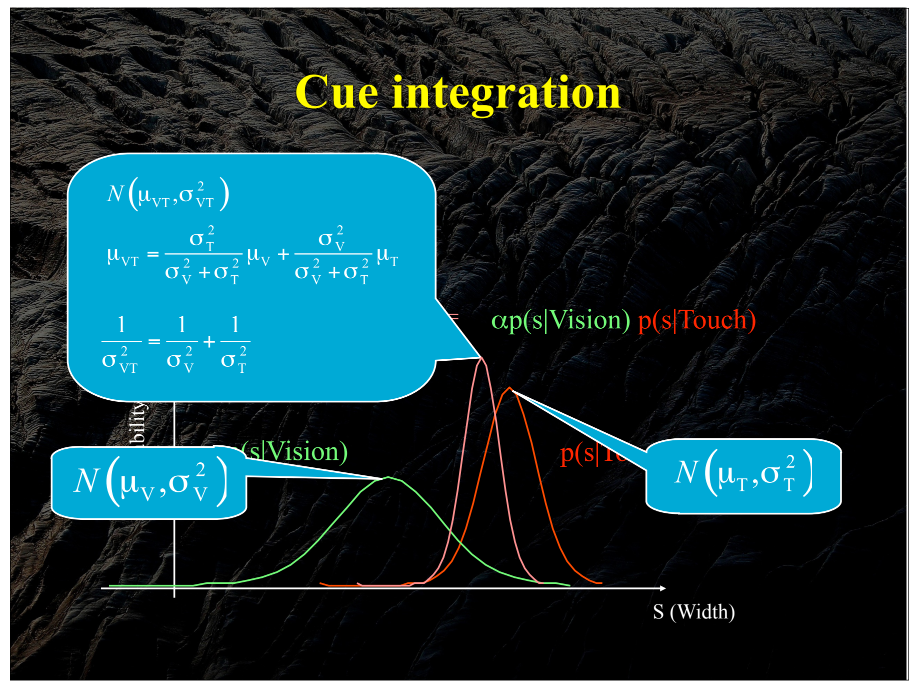
Bayesian approach: estimate the posterior p(s|r)

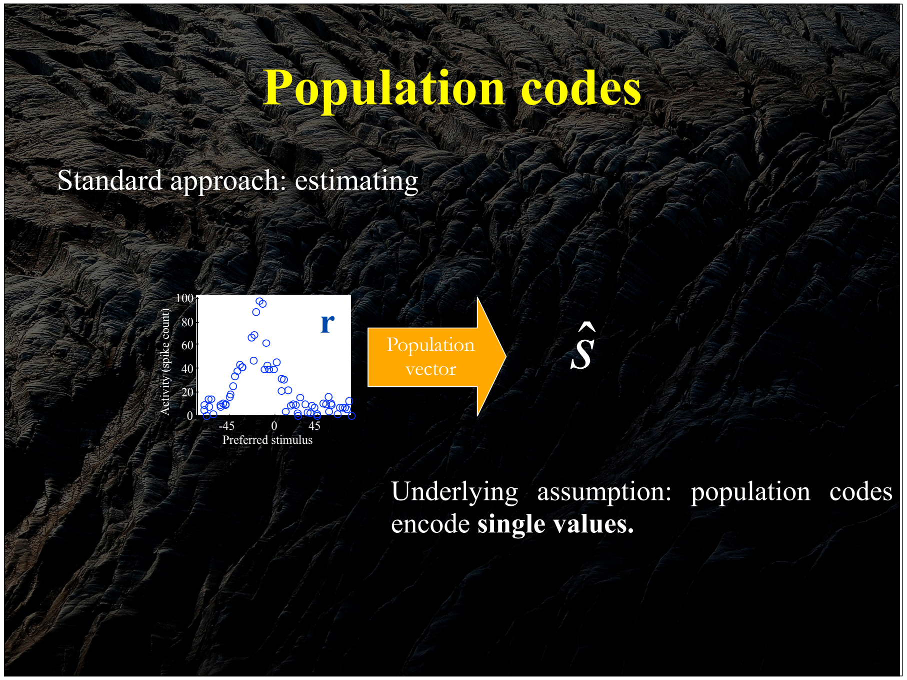
Advantages of a probabilistic representation

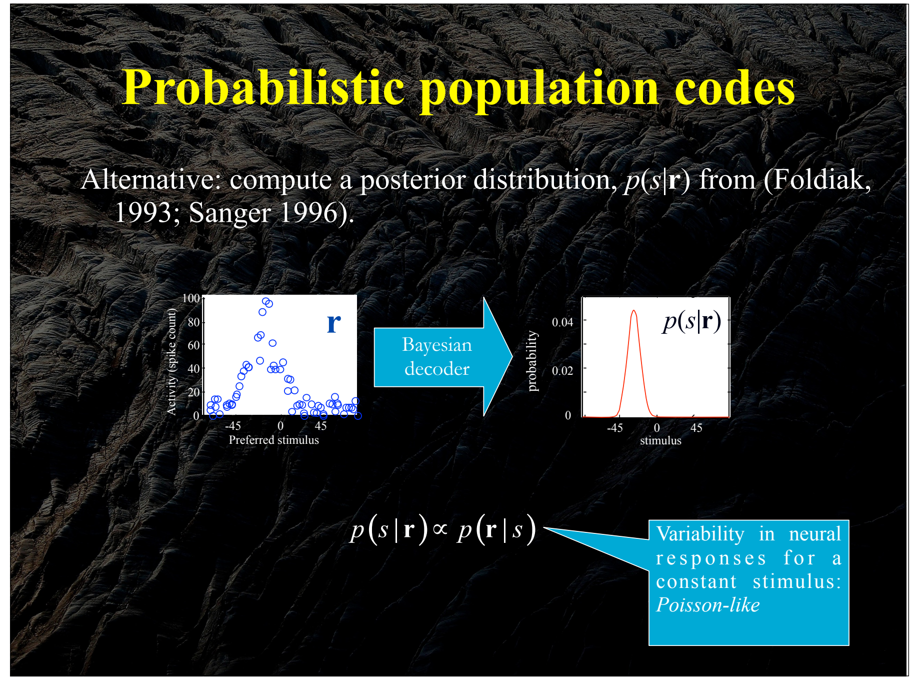


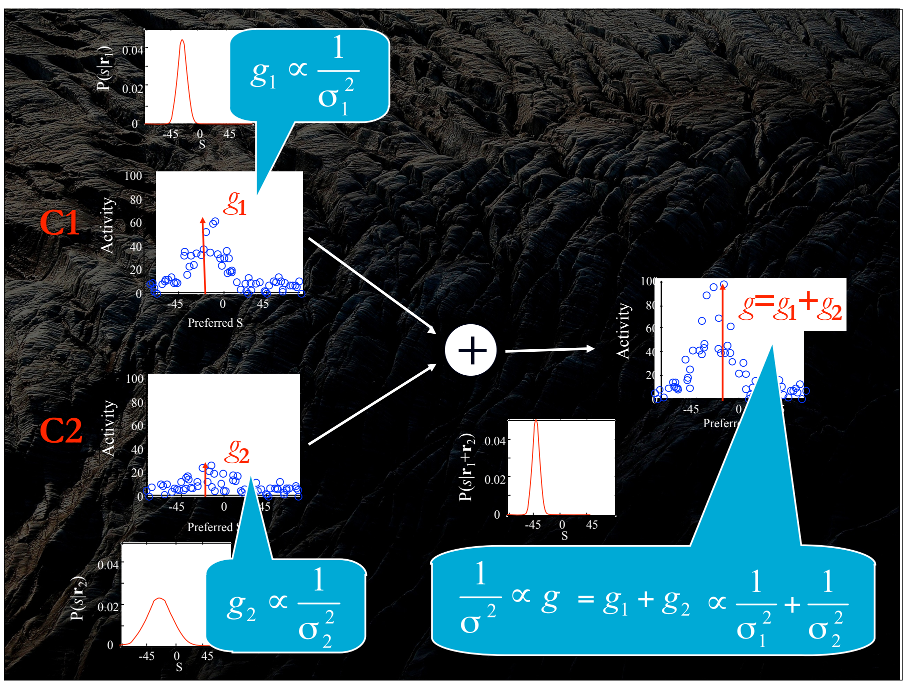
Recall Ex 3 in PS #3: Derive the optimal rule for integrating two noisy measurements to estimate the mean

$$\mu = \frac{r_1}{r_1 + r_2}\mu_1 + \frac{r_2}{r_1 + r_2}\mu_2$$




Population codes
Standard approach: estimating
Activity (spike count)
100
80
60
40
20
0
r
-45 0 45
Preferred stimulus
Population vector
ŝ
Underlying assumption: population codes encode **single values.**


Probabilistic population codes
Alternative: compute a posterior distribution, $p(s|\mathbf{r})$ from (Foldiak, 1993; Sanger 1996).
Activity (spike count)
r
-45 0 45
Preferred stimulus
Bayesian decoder
probability
0.04
0.02
0
$p(s|\mathbf{r})$
-45 0 45
stimulus
$p(s|\mathbf{r}) \propto p(\mathbf{r}|s)$
Variability in neural responses for a constant stimulus: *Poisson-like*


$P(s|\mathbf{r}_1)$
$g_1 \propto \frac{1}{\sigma_1^2}$
C1
Activity
g_1
Preferred S
C2
Activity
g_2
Preferred S
$P(s|\mathbf{r}_2)$
$g_2 \propto \frac{1}{\sigma_2^2}$
+
Activity
$g = g_1 + g_2$
$P(s|\mathbf{r}_1+\mathbf{r}_2)$
$\frac{1}{\sigma^2} \propto g = g_1 + g_2 \propto \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}$