Lecture 2

Non-Gaussian Statistics: Scene properties & models

Sounds

analogous work in domain of natural sounds (e.g., Attias & Schreiner, 1997)

examined low-order statistics of several sound ensembles (cat vocalizations, bird songs, wolf cries, environmental sounds, symphonic music, jazz, pop music, speech)

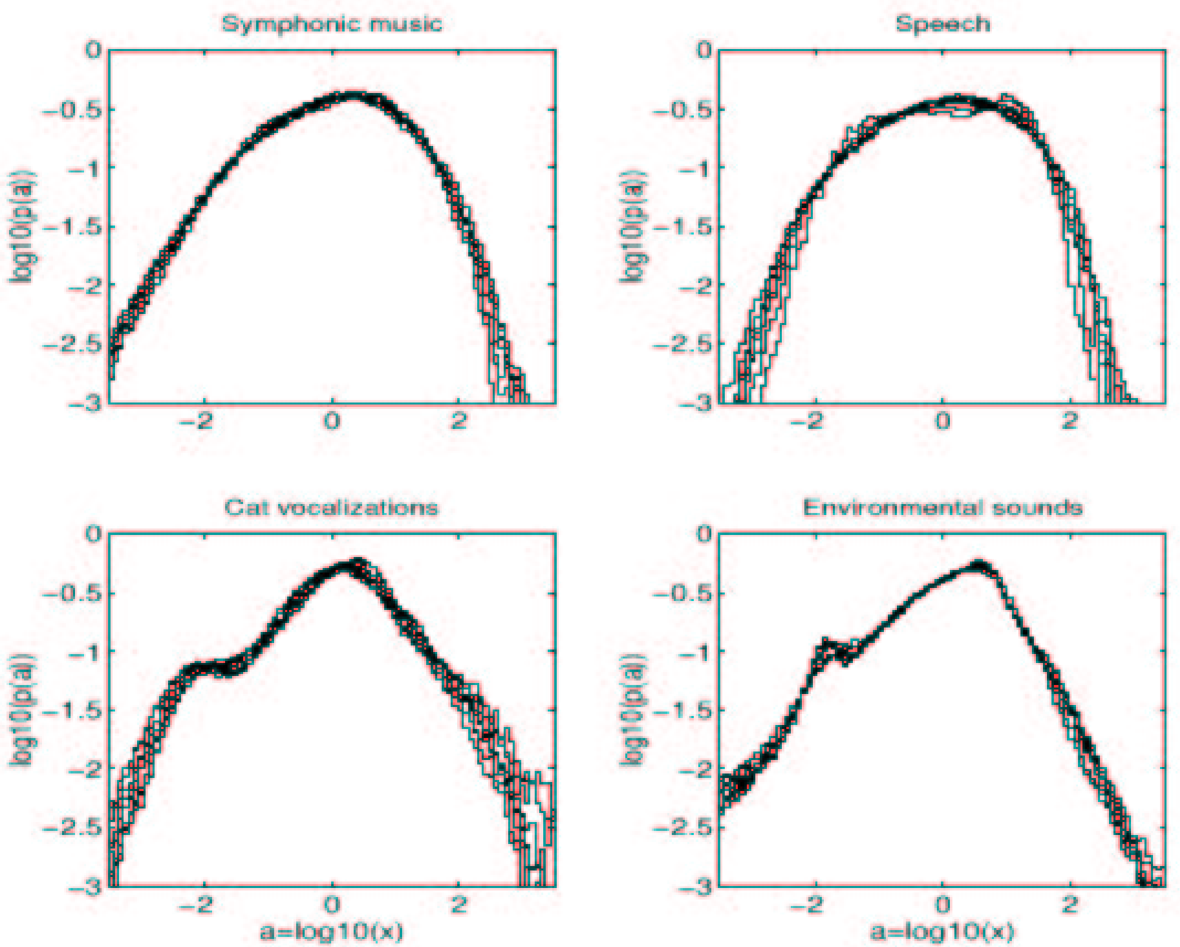
represent sound: 30 sec segments, sampled, represented in frequency bands – convolve with square non-overlapping filters, center frequencies $\nu = 100 - 11025$ Hz

focus on spectrotemporal amplitude (STA) $x_\nu(t)$:

$$s_\nu(t) = x_\nu(t) \cos(\nu t + \phi_\nu(t))$$

Sounds: Amplitude distribution

normalize amplitude distribution for given band (freq ν):
$< \log x_\nu(t) >= 0; < (\log x_\nu(t))^2 >= 1$



note:

- histograms for different bands agree
- exponential decay at high amplitudes
- long tail for low amplitudes (non-Gaussian) – abundance of soft sounds

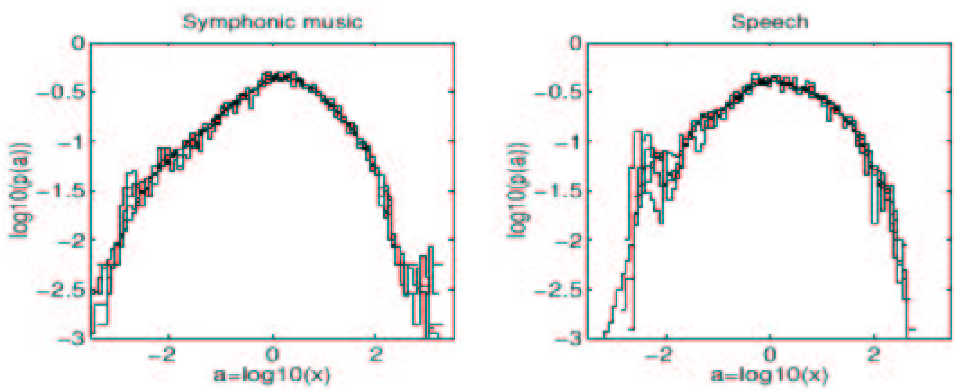
Sounds: Scale invariance

process is scale-invariant if any statistical quantity on given scale does not change as scale changed

look at different temporal resolutions:

$$x_{\nu}^{(n)}(t) = \frac{1}{n} \sum_{k=0}^{n-1} x_{\nu}(t + k/f_s)$$

histogram at $\nu = 800$Hz; $n = 1, 20, 50, 100, 200$: no central limit theorem



Relevance to sensory systems

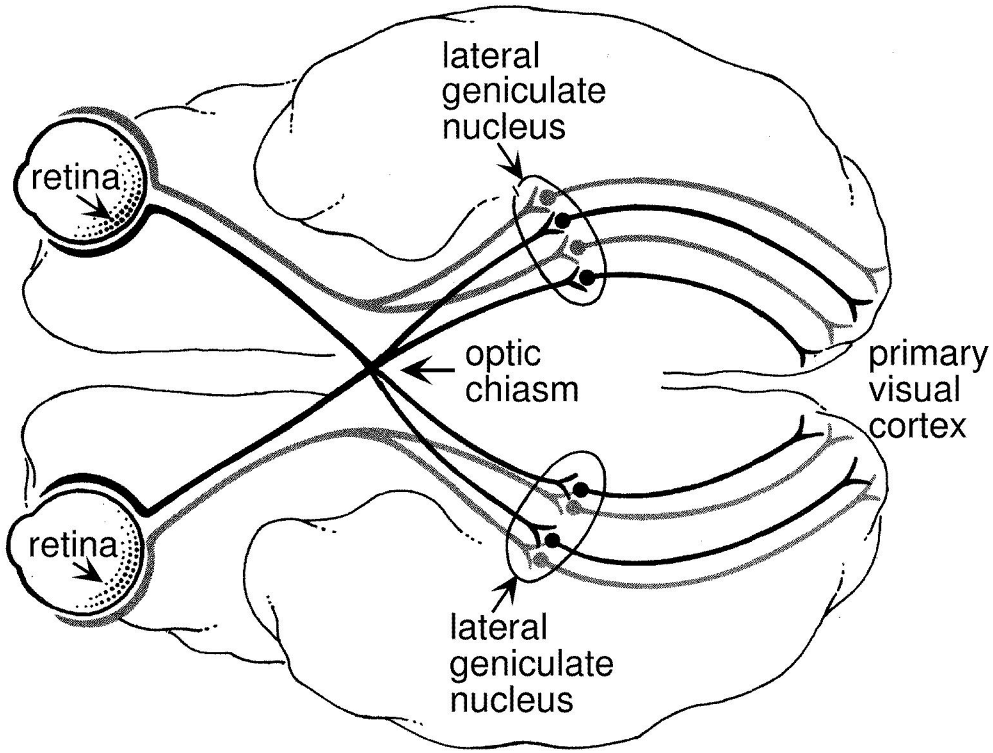
both natural sounds and images highly redundant

beneficial for auditory and visual systems to adapt representations to these statistics – improve discrimination ability

now look at early visual system, methods of characterizing cell responses

then relate to natural statistics

Early visual system

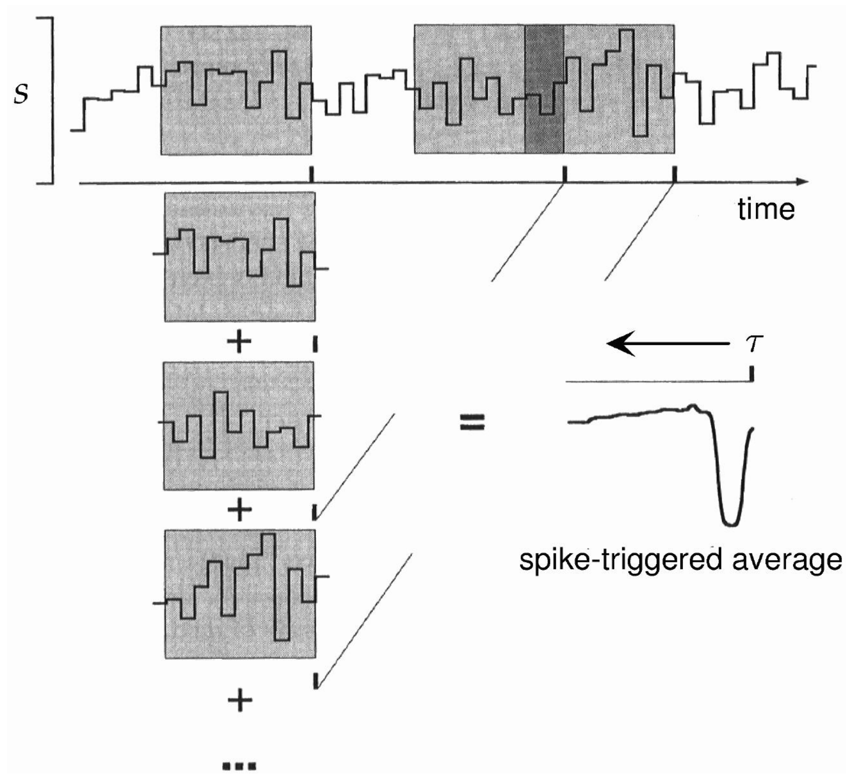


neurons in retina, LGN, and V1 (primary visual cortex) respond to light stimuli in restricted regions of visual field: receptive field (RF)

probed with spots, moving gratings – what causes cell to spike??

Spike triggered average

method of describing stimulus that causes cell to respond



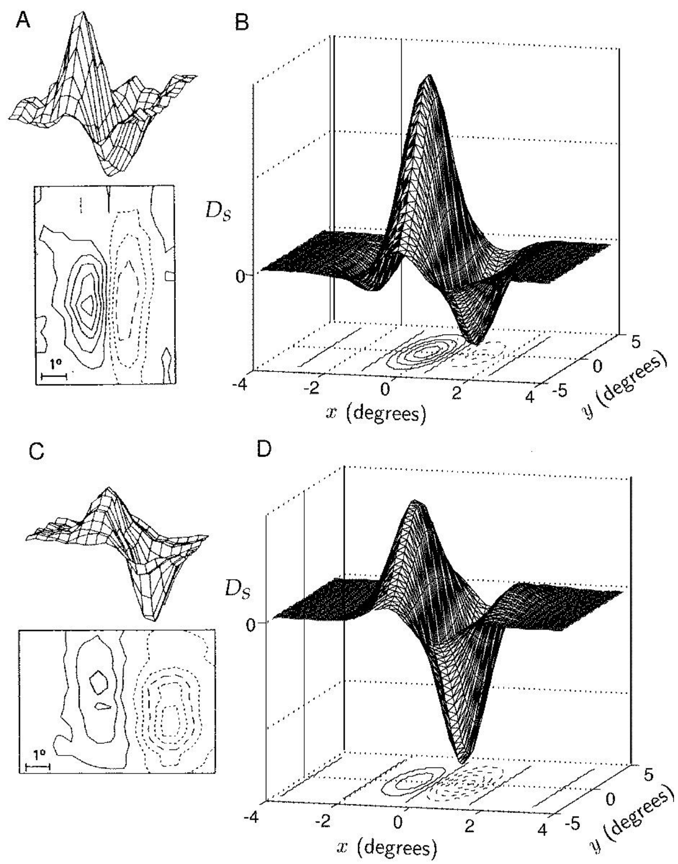
average over many spikes, many trials, stimuli

Simple cell receptive fields

STA $\rightarrow$ spatial RF structure for cat primary visual cortex

fit by Gabor functions (product of sinusoid and Gaussian):

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



Wavelets of Gabors

wide range of transforms capable of representing information in n dimensional data space (e.g., Fourier, Gabor)

wavelet: transform in which bandwidths increase proportionally to frequency; arrays of basis functions differ only by translations, dilations, and rotations of single function

wavelets based on Gabors are popular models of early visual cortex:

1. RFs localized in space, bandpass in frequency
2. frequency bandwidths constant when measured on log axes (octaves), so self-similar RFs (bandpass)
3. orientation selective (oriented)

Whoa

many properties of cortical simple cells not captured by this model:

1. end-stopping
2. cross-orientation inhibition
3. non-negative responses

only rough approximations to cells in visual cortex

Response to natural scenes

apply filters to natural images, examine statistics of responses

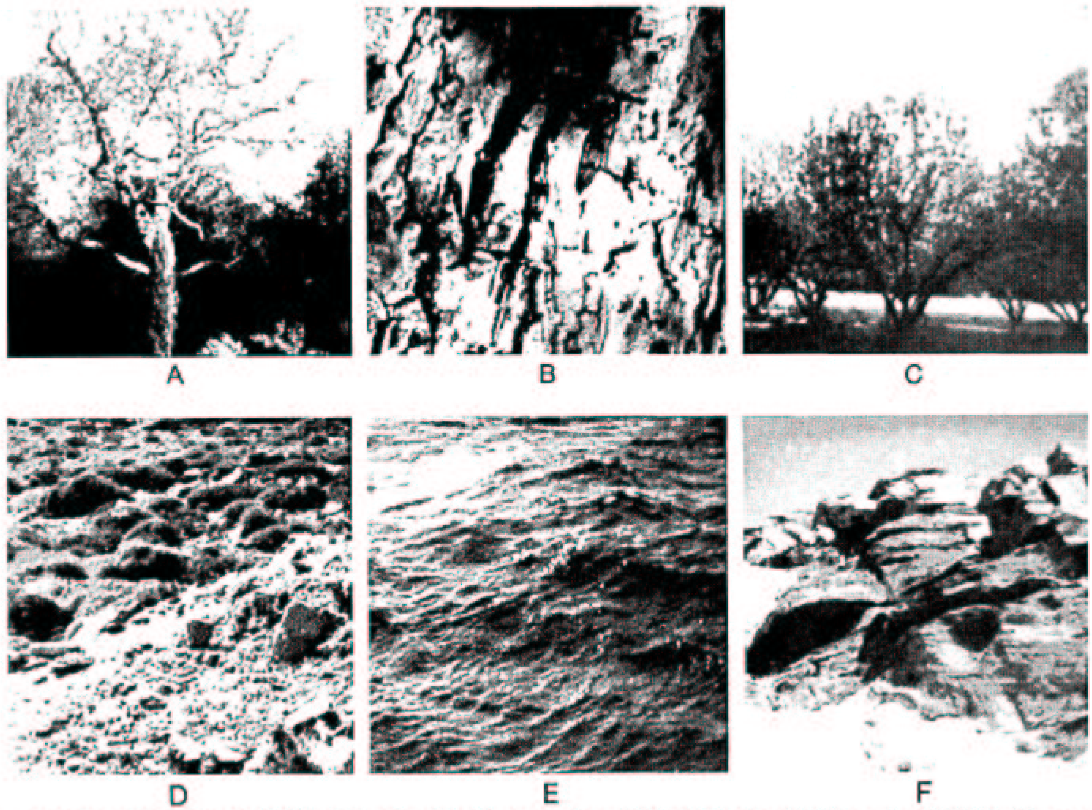
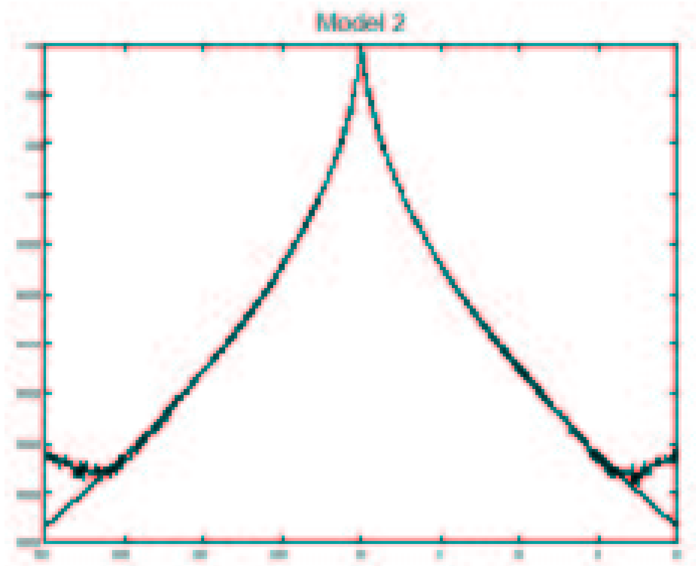


Fig. 6. Examples of the six images (A–F) in this study. Each image consists of 256 × 256 pixels with 256 gray levels (8 bits). However, only the central region was directly analyzed (160 × 160). See the text or details.

non-Gaussian responses – heavy tails, high prob of no response and large response relative to normal

Sparse responses

high probability of no response: sparseness (few of many possible units participate in coding of stimulus)



one property of distributions of sparse codes – high kurtosis

$$K = \frac{1}{n}\sum[(x-\mu)^4/\sigma^4] - 3$$

Reverse engineering: Efficient coding

Can filters be learned from images? Can we understand response properties of units in terms of strategy for processing natural images?

Barlow hypothesized that efficient coding of visual information is fundamental constraint on neural processing

maximize information that neural responses provide about visual environment

- responses of individual neurons to natural environment should fully utilize output capacity
- responses of different neurons to natural environment should be statistically independent of each other

translates into aim of reducing redundancy between neurons

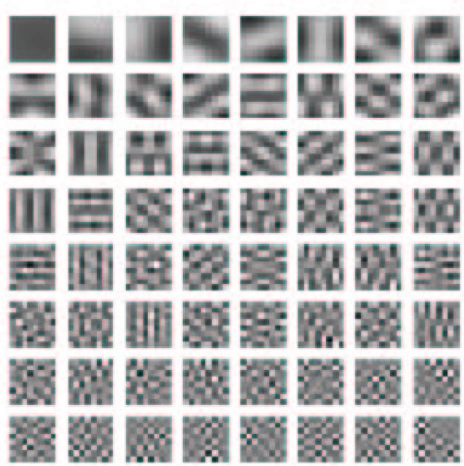
What filters reduce redundancy?

one proposal – Principal Components Analysis (PCA): computes eigenvectors of covariance matrix of data (e.g., covariance of pixels in image), produces orthogonal vectors, coefficients ordered by portion of covariance accounted for

retain top few vectors – minimal loss in data representation

removing low-probability regions reduces redundancy

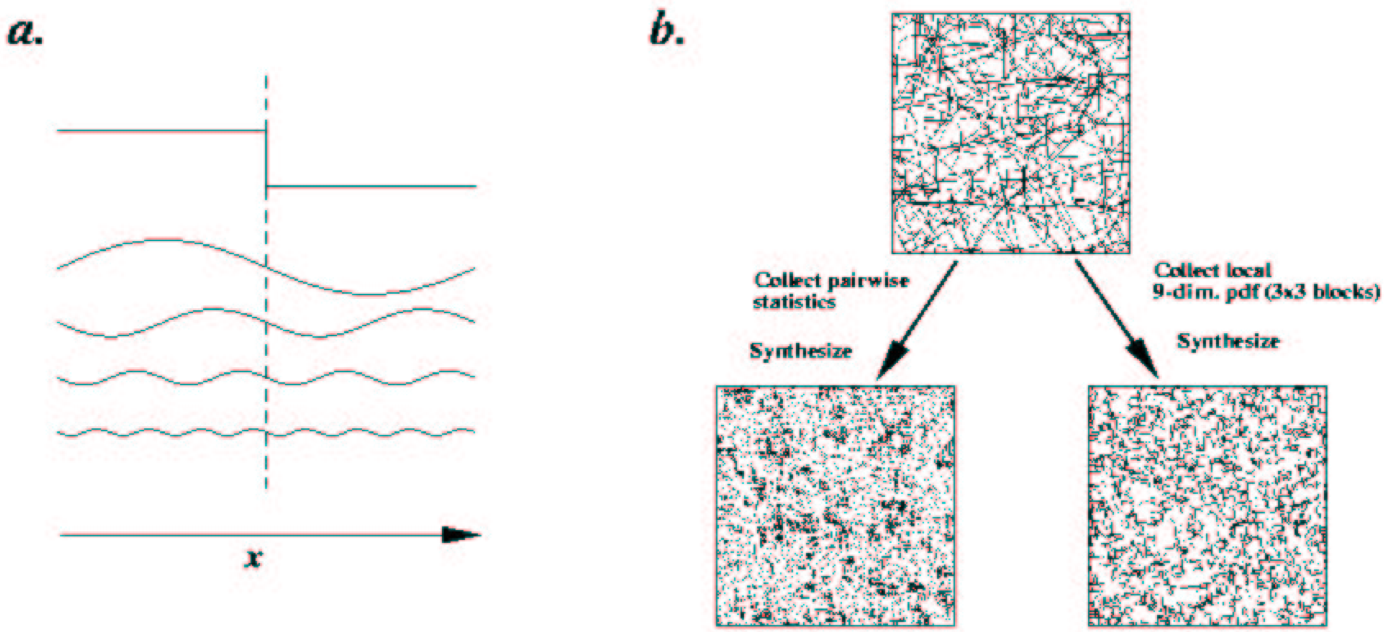
many similarities between principal components of natural scenes and RFs of visual cells, but not localized nor oriented



PCA inadequate

PCA based on covariance between pixels – only capable of learning pairwise correlations

pairwise correlations characterize only power spectrum, not phase alignment: cannot find phase alignment that occurs at edges, lines in images



instead try to learn simple filters (linear) that can still capture higher-order dependencies

Learning objective

hypothesize generative model of image $I(x,y)$:

$$I(x,y) = \sum_i a_i \phi_i(x,y)$$

$\phi_i()$ are filters, basis functions that form code for images; a_i are coefficients (filter responses)

objective or cost functional to minimize (gradient descent):

$$E(a,\phi) = \sum_{x,y} [I(x,y) - \sum_i a_i \phi_i(x,y)]^2 + \beta S(a_i/\sigma_i)$$

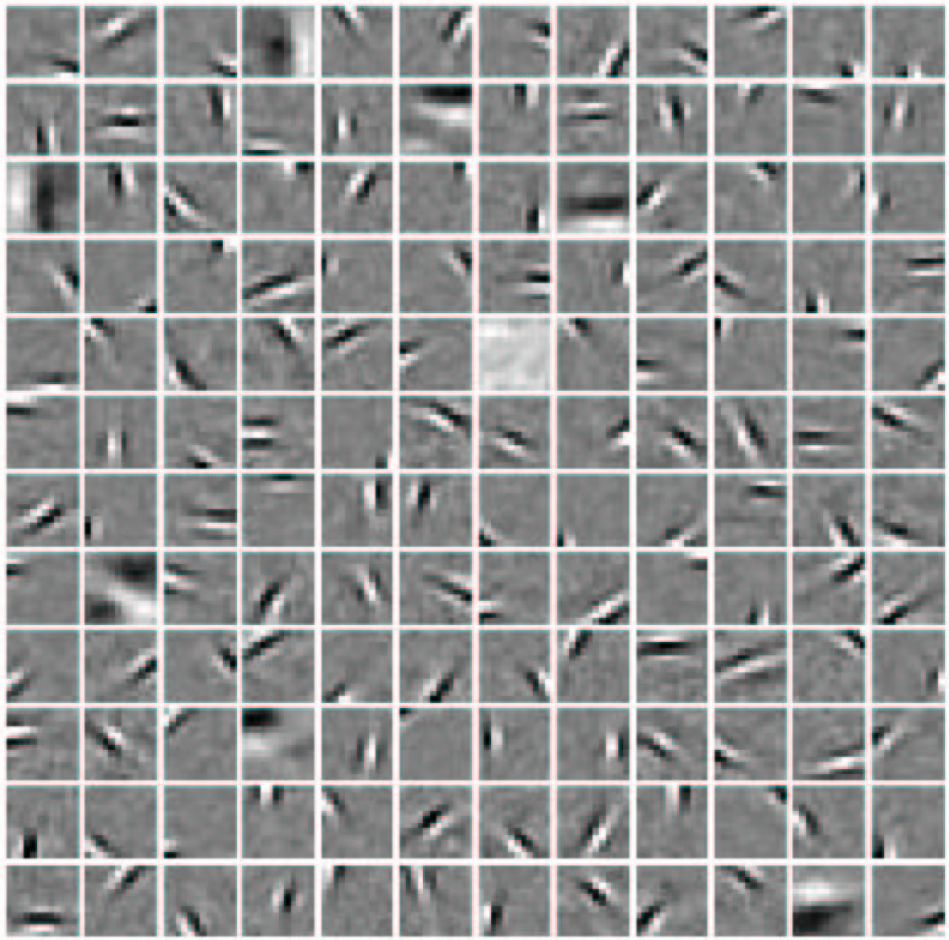
combines reconstruction cost with activity cost

expect a_i to be sparse, kurtotic, heavy-tailed, etc. $-$ log prior $S(x)$ can correspond to Cauchy ($\log(1+x^2)$); exponential ($|x|$); Laplacian

Results

train on 12x12 image patches extracted from natural scenes

learned filters are bandpass, localized, oriented:



but note these are projective fields, not receptive fields

Corresponding probabilistic model

choice of basis functions $\phi_i()$ determine image code:

$$I(\mathbf{x}) = \sum_i a_i \phi_i(\mathbf{x})$$

receptive fields determined by linear transform of image with other functions $\psi_i()$:

$$b_i = \sum_{\mathbf{x}_j} \psi_i(\mathbf{x}_j) I(\mathbf{x}_j) \mathbf{b} = \mathbf{W}\mathbf{I}$$

if ϕ linearly independent and same number as inputs, then $\phi_i(\mathbf{x}) = (\mathbf{W}^{-1})_{ji}$

if ϕ form orthonormal basis, then code is self-inverting: $\phi_i(\mathbf{x}) = \psi_i(\mathbf{x})$

image model is over-complete if more basis functions than effective dimensions of input