I-tutorial
Learning of Invariant Representations in Sensory Cortex

tomaso poggio
CBMM
McGovern Institute, BCS,
LCSL, CSAIL
MIT
1. Intro and background

2. Mathematics of invariance

3. Biophysical mechanisms for tuning and pooling

4. Retina and V1: eccentricity dependent RFs; V2 and V4: pooling, crowding and clutter

5. IT: Class-specific approximate invariance and remarks
Class 23
Wed Nov 26
Learning Invariant Representations:
3. Biophysical mechanisms for tuning and pooling
Summary of previous class
M-Theory

So far: compact groups in $R^2$

M-theory extends proves invariance + uniqueness theorems for

- partially observable groups
- non-group transformations
- hierarchies of magic HW modules (multilayer)
End Summary
Biophysical mechanisms for tuning and pooling

- A single cell model of simple and complex cells (any fixed nonlinearity is OK)
- Hebb synapses, PCAs and Gabor
- PCAs, Foldiak and pooling
- Mirror-symmetric tuning in the face patches
Our basic machine: a HW module
(dot products and histograms for an image in a receptive field window)

- The cumulative histogram (empirical cdf) can be be computed as

\[ \mu^k(I) = \frac{1}{|G|} \sum_{i=1}^{G} \sigma(\langle I, g^k_i \rangle + n\Delta) \]

- This maps directly into a set of simple cells with threshold

- …and a complex cell indexed by n and k summatting the simple cells
Dendrites of a complex cells as *simple cells*…
Biophysical mechanisms for tuning and pooling

- A single cell model of simple and complex cells (any fixed nonlinearity is OK)
- Hebb synapses, PCAs and Gabor
- Pooling
- Mirror-symmetric tuning in the face patches
Linking Conjecture

- The memory in a layer of cells (such as simple cells in V1) is stored in the weights of the connections between the neurons and the inputs (from the previous layers).
- Instead of storing a sequence of discrete frames (the templatebook) as assumed in Part I, online learning is more likely, with synaptic weights being incrementally modified during development.
- Hebbian-like synapses exist in visual cortex.
- Hebbian-like learning is equivalent to an online algorithm computing PCAs.
- As a consequence, the tuning of simple cortical cells is dictated by the PCAs of the templatebook.
Hebb synapses imply that the tuning of the neuron converges to the top eigenvector of the covariance matrix of the “frames” of the movie of objects transforming. The convergence follows the Oja flow

\[ t_{k+1} - t_k = x \cdot y + n(t, y) \]

Different cells are exposed (during development) to translations in different directions.
Gaussian aperture: the cortical equation

Define as templatebook $T$ the matrix where each column represents a template $t$ shifted relative to the previous column and “seen through a Gaussian aperture”. The image is assumed to be 1D. The image seen through a Gaussian aperture is then $t(y - x)g(x)$ when the image is shifted by $y$. We are led to the following problem: find the eigenvectors of the symmetric matrix $G^T T^T T G$ where $G$ is a diagonal matrix with the values of a Gaussian along the diagonal. We consider the continuous version of the problem, that is the eigenvalue problem

$$\int dxg(y)g(x)\psi_n(x) \int ds\bar{t}(y - s)\bar{t}(s - x) = \lambda_n \psi_n(y)$$

which is rewritten as the cortical equation

$$\int dxg(y)g(x)t(y - x)\psi_n(x) = \lambda_n \psi_n(y)$$

with $t(x)$ being the autocorrelation function of the template. This is an equation describing the development of simple cells in V1; it describes development of other cortical layers as well.
The cortical equation: general properties

The equation

$$\int dx g(y)g(x)\psi_n(x)t(y-x) = \lambda_n \psi_n(y)$$

(3)

with $t(x)$ being real and even, is an eigenvalue equation for an integral operator with symmetric positive definite kernel $K(x, y) = g(y)g(x)t(y-x)$, $K(x, y) = K(y, x)$. The kernel is measurable on the square $[a, b] \times [a, b]$, since

$$\int_a^b \int_a^b |K(x, y)|^2 dx dy < \infty.$$  (4)

Then the Hilbert-Schmidt theory of symmetric kernels applied to (3) implies

- the eigenfunctions corresponding to distinct eigenvalues are orthogonal
- the eigenvalues are real and positive
- there is at least one eigenvalues and one eigenfunctions (when $K$ is almost everywhere nonzero) and in general a countable set of eigenfunctions.
The cortical equation: approximate solution: Gabor wavelets!

We represent the template as \( t(x) = \int d\omega \ t(\omega) e^{i\omega x} \) and assume that the eigenfunction has the form \( \psi(x) = e^{-\frac{\beta}{2} x^2} e^{i\omega_g x} \), where \( \beta \) and \( \omega_g \) are parameters to be found. We have that the cortical equation becomes

\[
e^{-\frac{\alpha}{2} y^2} \int dx \ e^{-\frac{x^2(\alpha+\beta)}{2}} \int d\omega \ t(\omega) e^{i\omega(y-x)} e^{-i\omega_g x} = \lambda(\omega) e^{-\frac{\beta y^2}{2}} e^{-i\omega_g y}
\]

Collecting the terms in \( x \) and integrating we have that the r.h.s become:

\[
\sqrt{\frac{2\pi}{\alpha + \beta}} e^{-\frac{\alpha}{2} y^2} \int d\omega \ t(\omega) e^{-i\omega y} e^{-\frac{2(\omega_g - \omega)^2}{\alpha + \beta}}
\]

With the variable change \( \tilde{\omega} = \omega_g - \omega \) and in the hypothesis that \( t(\tilde{\omega}) \approx \text{const} \) for a few standard deviations of the Gaussian centered in \( \bar{\omega} \), that is \( \pm 3\sqrt{(\alpha + \beta)/2} \) of \( \omega \sim \omega_g \), integrating in \( \tilde{\omega} \) we have:

\[
e^{-\frac{\nu^2 \alpha}{2}} e^{-i\omega_g y} e^{-\frac{\nu^2(\alpha+\beta)}{2}} \sim \lambda(\omega_g) e^{-\frac{\nu^2 \beta}{2}} e^{-i\omega_g y}.
\]

Notice that this implies an upper bound on \( \beta \) since otherwise \( t \) would be white noise which is inconsistent with the diffraction-limited optics of the eye.
Cortical equation in 2D: natural images, Gabor-like receptive fields

Rust et al. 2005

Carandini

Rust et al. 2005
In 1D the eigenvectors are Gabor like functions. In 2D the solutions are also Gabor with an orientation orthogonal to the direction of motion. Motion, together with high-pass filtering in the retina induces symmetry breaking that allows non-symmetric solution to emerge. Note that for motion at constant speed

\[
\frac{d}{dt} = \nu \frac{d}{dx}
\]
Figure 4: Retinal processing pipeline used for V1 simulations.
Gaussian aperture: the cortical equation

\[ \int dx g(y)g(x)t(y-x)\psi_n(x) = \lambda_n \psi_n(y). \]

with \( t(x) \) being the autocorrelation function of the template.

This is an equation describing the development of simple cells in V1; it describes development of other cortical layers as well.
Prediction agrees with data!
Cortical equation in 2D: natural images, Gabor-like receptive fields

Rust et al. 2005

Carandini

Rust et al. 2005
Cortical equation in 2D: natural images, Gabor-like receptive fields
Cortical equation in 2D: natural images, Gabor-like receptive fields
Natural images, Gabor-like receptive fields
Natural images, Gabor-like receptive fields
Natural images, Gabor-like receptive fields
Prediction agrees with data (not fully)
Beyond V1: work in progress

In the first layer exposure to translations determines the development of a set of receptive fields which are an overcomplete set of Gabor-like wavelets. The space of two-dimensional images – functions of $x, y$ – is effectively expanded into a 4-cube of wavelets where the dimensions are $x, y, \theta, s$, eg space, orientation and scale. The same online learning at the level of the second layer (S2) with apertures “looking” at a Gaussian ball in $x, y, \theta, s$ will converge to Gabor-like wavelet after exposure to image translations, which induce translations in $x, y$ of the 4-cube. In this case, the second-order wavelets are wavelets parallel to the $x, y$ plane of the 4-cube. For image motion that include rotations and looming, the resulting motion in the 4-cube is mostly still locally a shift – but in general along a diagonal in the 4-cube. Thus, in general, second-order wavelets are Gabor-like oriented along diagonals in $x, y, \theta, s$. 
Beyond V1, towards V2 and V4
Beyond V1, towards V2 and V4: wavelets of wavelets
We are working on implementing the full theory (the corresponding model contains Hmax and convolutional networks as special, simple cases)
Higher layers general properties

Solving the spectral problem:

\[ \mathbf{T}_t^* \mathbf{T}_t \psi_i = \lambda_i \psi_i \]

we have the following two results:

- The number of eigenfunctions depends on the size of the receptive field
- The eigenvalues obey the relation:

\[ \frac{\lambda_i(\sigma)}{\lambda_i(\bar{\sigma})} \geq 1, \quad \sigma \geq \bar{\sigma} \]

(Giacomo Spigler)
Cortical equation in 2D: natural images, Gabor-like receptive fields

Rust et al. 2005

Carandini

Rust et al. 2005
Prediction agrees with data (not fully)
V2 and V4: wavelets of wavelets

Responses of two model complex cells pooling 3D wavelets (top) and two real V4 cells (bottom) to various stimuli used by Gallant.

Red/orange indicates a high response and blue/green indicates a low response.
Face patches in IT
Face patches
Magic algorithm “works”
Train once, test everywhere

- Takes in two unaligned, non-detected face images.

![Mona Lisa](image1.jpg) ![Man](image2.jpg)

**Same person?**

- Go through the detection, alignment and recognition pipeline.

- Fully unconstrained face verification (as long as the face can be detected by the detector with relative good accuracy, usually we get > 80% verification accuracy.)
Even with the very simple classifier, our system’s performance still compares favorably with the current state of the art. In the case of LFW, our model’s performance exceeds the current state-of-the-art for an unsupervised system (86.2% using LQP — Local Quantized Patterns [30]—Note: ing), though the best supervised systems do better⁷. The strongest result in the literature for face verification with PubFig83¹⁰ is 70.2% [4]—which is 6.2% lower than our best model.
Linking Conjecture

• Predicts Gabor-like tuning of simple cells in V1

• Qualitatively predicts tuning in V2/V4

• Predicts/justifies mirror-symmetric tuning of cells in face patch AL
Magic architecture for pose-invariant face identification

Viewpoint tolerant units (complex units)

View-tuned units, tuned to full-face templates for different view angles
If Hebbian learning holds also for the face patches, then the theory predicts that between the view tuned patches and the view-invariant patch there should be mirror symmetric cells.

Lemma: PCAs for faces are odd or even functions, and so energy models complex cells are always even.
Response of simple AL “model” cells to different views of a face of brain theory
Remarks
The ventral stream:
a
architect that learns and discount transformations
architect determined by invariances of the physical world

• The goal of the cortical hierarchy is learning and factoring out transformations (*learning from data “any” invariance*)

• Features and statistics of natural images do not matter, transformations do, are learnable from random noise images

• Most important are complex cells and their pooling domain: this represents symmetries/invariances learned from the world -- eg equivalence classes of templates

• Invariances are learned by memory-based computation, just recording

• Aperture size determines which transformation, transformation determines tuning in V1 V2 V4

• Architecture predicts stability of visual perception wrt small eye motions

• Class-specific transformations predict class-specific modules
Higher layers general properties

Solving the spectral problem:

\[ T_t^* T_t v_i = \lambda_i v_i \]

we have the following two results:

- The number of eigenfunctions depends on the size of the receptive field
- The eigenvalues obey the relation:

\[ \frac{\lambda_i(\sigma)}{\lambda_i(\bar{\sigma})} \geq 1, \quad \sigma \geq \bar{\sigma} \]

(Giacomo Spigler)
Top module

\[ \Sigma = \text{signature vector} \]

Associative memory