Normative Modelling of the Visual System


- Normative vs Descriptive Theories: how *should* the system behave?
- Of course, this makes most sense if evolution has optimized the natural system. Effect of constraints
- “Statistical-ecological” approach

Chapter 10 of Dayan and Abbott (2001) is also useful.

Statistical-ecological approach

(HHH, p 21)

- Different sets of features are good for different kinds of data.
- The images that our eyes receive have certain statistical properties (regularities).
- The visual system has learned a model of these statistical properties.
- The model of the statistical properties enables (close to) optimal statistical inference.
- The model of the statistical properties is reflected in the measurable properties of the visual system (e.g. receptive fields of the neurons)

Redundancy Reduction

Ganglion cells: Whitening

Retina

LGN

V1

Simple cells: Independent Component Analysis (ICA)

Complex cells: Subspace ICA


Bandpass filtering

Orientation selectivity

Phase invariance

Slide credit: Matthias Bethge
Mutual Information and Populations of Neurons

\[ H(R) = -\int p(r) \log_2 p(r) \, dr - N \log_2 \Delta r \]

and

\[ H(R_a) = -\int p(r_a) \log_2 p(r_a) \, dr - \log_2 \Delta r \]

We have

\[ H(R) \leq \sum_a H(R_a) \]

(proof, consider KL divergence)

Recall that

\[ I(R; S) = H(R) - H(R|S) \]

so if noise entropy \( H(R|S) \) is independent of the transformation \( S \to R \), we can maximize mutual information by maximizing \( H(R) \) under given constraints.

Factorial Coding

- Maximization of population response entropy is achieved by
  1. factorial coding \( p(r) = \prod_a p(r_a) \)
  2. each response distribution must be optimized wrt the imposed constraints
- If all neurons have the same constraints \( \Rightarrow \) probability equalization. This does not mean that each variable responds identically!
- Exact factorization and probability equalization are difficult to achieve
- A more modest goal is decorrelation (whitening)
  \[ \langle (r - \langle r \rangle)(r - \langle r \rangle)^T \rangle = \sigma_r^2 \mathbf{I} \]

Second order statistics

- First order image statistics \( \langle s(x, t) \rangle \)
- Second order, correlation \( Q(x, x', t, t') = \langle s(x, t)s(x', t') \rangle \)
- By Wiener-Kinchin specifying \( Q \) is equivalent to specifying \( PSD = |\tilde{s}(f)|^2 \) (Wiener-Kinchin)
- Gaussian approximation \( \Leftrightarrow Q(x, x') \Leftrightarrow PSD \)
- Higher order statistics, e.g. \( \langle s(x, t)s(x', t')s(x'', t'') \rangle \)
  will be discussed later

Principal Component Analysis

- Subtract mean of \( s \)
- Linear model (!): \( r = Ws \)
- One solution for \( W \): PCA. Find the eigenvectors of \( \text{cov}(s) = \langle ss^T \rangle = Q_{ss} \) and scale
- Write \( Q_{ss} = U \Lambda U^T \) (where \( U^T U = I \) and \( \Lambda \) is diagonal). Set \( W = \Lambda^{-1/2} U^T \), then \( \langle rr^T \rangle = I \)
- First PC maximizes \( \text{var}(w_1 \cdot s) \) subject to \( |w_1|^2 = 1 \)
- Subsequent components: subtract previous ones and repeat procedure
- Can be used for dimensionality reduction by removing modes with lowest eigenvalues.
PCA on Natural Image Patches

If translation invariant covariance matrix, $C_{ij} = f(|i-j|)$:
eigenvectors are periodic (proof: e.g. HHH p.125).
So PCA = Fourier analysis.

Whitening with PCA

To whiten: 1) do PCA projections 2) scale components with inverse variance.

Generative model with PCA

$$s = \sum_k w_k r_k$$
$$P(r) = \prod_k P(r_k) = \prod_k N(0, \sigma_k^2)$$
Gaussian mix of principal components

Importance of Fourier Phase Information

Left: sample images.
Right: a) phase of (a) + amplitude of (b), b) v.v.
(Method: Fourier transform image, split into magnitude and phase, mix, inverse transform)

- PSD contains no phase information, so second order stats miss important information ... tbc.
PCA is not the only solution to whitening

\[ r = Ws. \] For many choices of \( W \), \( \text{cov}(r) \) is diagonal.

- PCA (see above)
- ZCA (Z for zero-phase, as filters are symmetrical. Set \( W = Q_{ss}^{-1/2} \). [Bell and Sejnowski, 1997].
- Note that for \( \text{cov}(r) = I \), if \( W \) is a solution then so is \( W' = RW \), where \( R \) is a rotation matrix (so \( R^T = R^{-1} \))

See also Hyvärinen, Hurri and Hoyer (2009) §5.9

Cov function \( k(x, x') = \exp(-|x - x'|/\ell) \) has power spectrum
\[ \propto k^{-2} \] asymptotically

\[ Q_{ss} \]
\[ W_{ZCA} = Q_{ss}^{-1/2} \]

One row of \( W_{ZCA} \) 2-D whitening filter (HHH)

W - space

LOCAL IN FREQUENCY
GLOBAL IN SPACE
LOCAL IN SPACE
GLOBAL IN FREQUENCY

PCA

ZCA

ICA

SEMI-LOCAL

Decorrelating W’s: \( W^TW = \langle x x^T \rangle^{-1} \)

[Bell and Sejnowski, 1997]
Retinal Ganglion Cell Receptive Fields

Continuous-space version of the above calculation
See [Atick and Redlich, 1990], we consider spatial part of the calculation only. See also Dayan and Abbott §4.2

\[
\begin{align*}
\mathbf{r}(a) &= \int D(x - a)s(x)dx \\
Q_{rr}(a, b) &= \int \int D(x - a)D(y - b)\langle s(x)s(y) \rangle dx dy
\end{align*}
\]

For decorrelation we require

\[
Q_{rr}(a, b) = \sigma^2_r\delta(a, b)
\]

Do calculations in the Fourier basis

\[
\tilde{D}(\kappa) = \int D(x)\exp(i\kappa \cdot x)dx
\]

\[
D(x) = \frac{1}{4\pi^2} \int \tilde{D}(\kappa)\exp(-i\kappa \cdot x)d\kappa
\]

Filtering Input Noise

Total input is \(s(x) + \eta(x)\), where \(\eta(x)\) is noise, reflecting image distortion, photoreceptor noise etc

Optimal least-squares filter is the Wiener filter with

\[
\tilde{D}_\eta(\kappa) = \frac{\tilde{Q}_{ss}(\kappa)}{\tilde{Q}_{ss}(\kappa) + \tilde{Q}_{\eta\eta}(\kappa)}
\]

Thus

\[
\tilde{D}_s(\kappa) = \tilde{D}(\kappa)\tilde{D}_\eta(\kappa)
\]

\[
|\tilde{D}_s(\kappa)| = \frac{\sigma_r}{\sqrt{\tilde{Q}_{ss}(\kappa) + \tilde{Q}_{\eta\eta}(\kappa)}}
\]
For low noise the kernel has a bandpass character, and the predicted receptive field has a centre-surround structure.

This eliminates one major source of redundancy arising from strong similarity of neighbouring inputs.

For high noise the structure of the optimal filter is low-pass, and the RF loses its surround.

This averages over neighbouring inputs to extract the signal which is obscured by noise.

Result is not simple PCA as we have enforced spatial invariance on the filter.

In the retina, low light levels $\equiv$ high noise. The predicted change matches observations [Van Nes and Bouman, 1967].
Contribution of Spiking to de-correlation

[Pitkow and Meister, 2012]

Decorrelation from Centre-Surround RF

but spiking threshold contributes even more.

Further Decorrelation Analyses

- Spatio-temporal coding (Dong and Atick, 1995; Li, 1996). Power spectrum is $1/f^2$ but non-separable
- Colour opponency: red centre, green surround (and vice versa) [Atick et al., 1993]

Caveats for the Information Maximization Approach

- Information maximization sets limited goals and requires strong assumptions
- Analyzes representational properties but ignores computational goals e.g. object recognition, target tracking
- Cortical processing of visual signals requires analysis beyond information transfer. V1 can have no more information about the visual signal than the LGN, but it has many more neurons
- However, information transfer analysis does help understand mutual selectivities: RFs with preference for high spatial frequencies are low-pass temporal filters, and RFs with selectivity for low spatial frequency act as bandpass temporal filters

References I
