A metric space approach to the information capacity of spike trains.

Conor Houghton

Mathematical Neuroscience Laboratory
School of Mathematics
Trinity College Dublin

Berlin, 30 March 2011
A metric space approach to the information capacity of spike trains.
Questions about trains.

- How do the properties of spike trains change along a sensory pathway?
- Is the song rate coded or is there information in temporal features?
- Are neurons in populations redundant?
Questions about trains.

- What is the information theory of spike trains?
A metric space approach to the information capacity of spike trains.

Spike trains.

Information theory.
A metric space approach to the information capacity of spike trains.

— Spike trains.

Shannon’s entropy.

\[ H = - \sum_{\text{events}} (\text{probability of the event}) \log (\text{probability of the event}) \]
Bialek approach to information and spike trains.¹

Bialek approach to information - calculation.

Make a table, for example:

<table>
<thead>
<tr>
<th>word</th>
<th>00101</th>
<th>01001</th>
<th>10110</th>
<th>10101</th>
<th>10110</th>
<th>10000</th>
<th>etc</th>
</tr>
</thead>
<tbody>
<tr>
<td>prob.</td>
<td>0.011</td>
<td>0.022</td>
<td>0.052</td>
<td>0.011</td>
<td>0.054</td>
<td>0.098</td>
<td>…</td>
</tr>
</tbody>
</table>

and calculate the corresponding entropy

\[ H = - \sum_{\text{words}} p(\text{word}) \log p(\text{word}) \]
Bialek approach to information - result.

The information is the difference between the signal entropy and the noise entropy.

\[ H_s - H_\eta \]
This is the *mutual information* between the stimulus and the response.

\[
H_s = H(\text{response}) \\
H_\eta = H(\text{response} | \text{stimulus})
\]

so

\[
I(\text{response, stimulus}) = H_s - H_\eta
\]
Bialek approach to information and spike trains - problems.

- To take into account timing precision a small discretization lengths is needed.
- A huge number of words, most of the ones that occur are mostly zeros.
- A huge sample size needed; Bialek worked with fly, such long recording are not normally possible.
- There are also interpretational problems with any information theory approach to neuroscience, we won’t deal with that here.
Bialek approach to information and spike trains - no noise model.

- No model of noise.
- No notion of one word being near another.
A metric space approach to the information capacity of spike trains.

The space of spike trains?

• Should we using the discrete theory or the continuous one?
• Spike times are not discrete.

• The continuous theory assumes a continuous space, what is the space of spike trains?
A metric space approach to the information capacity of spike trains.

Metric spaces

A metric maps pairs of points \( a \) and \( b \), to a real number \( d(a, b) \) such that

- Positive and distinguishable
  \[
  d(a, b) \geq 0 \\
  d(a, b) = 0 \iff a = b,
  \]

- Symmetric
  \[
  d(a, b) = d(b, a).
  \]

- Triangle inequality
  \[
  d(a, b) \leq d(a, c) + d(c, b).
  \]
Euclidean metrics

- In $\mathbb{R}^3$ say
  
  \[
  \mathbf{x} = (x_1, x_2, x_3) \\
  \mathbf{y} = (y_1, y_2, y_3)
  \]

- The dot product is given by
  \[
  \mathbf{x} \cdot \mathbf{y} = x_1 y_1 + x_2 y_2 + x_3 y_3
  \]

- The dot-product of a vector with itself is a norm, a measure of the length of the vector $|\mathbf{x}| = \sqrt{\mathbf{x} \cdot \mathbf{x}}$.

- This norm induces a metric, called the $L^2$ metric

  \[
  d(\mathbf{x}, \mathbf{y}) = |\mathbf{x} - \mathbf{y}| = \sqrt{\sum_{i=1}^{3} (x_i - y_i)^2}.
  \]
A metric space approach to the information capacity of spike trains.

Metric spaces.

Euclidean metrics on the space of functions.

This generalizes to functions, if \( f(t) \) and \( g(t) \) are both real functions on the same interval, \([0, T]\) say, then the \( L^2 \)-metric is

\[
d(f, g) = \sqrt{\int_0^T dt(f - g)^2}.
\]
Spike trains aren’t a vector space.

• While it might be possible to define the addition of two spike trains by superposition, it isn’t at all obvious how to define the difference.
• There is no reason to expect spike trains to be Euclidean.
A metric space approach to the information capacity of spike trains.

A non-Euclidean metric: Metrics in towns.

‘As the crow flies’ distance versus route distance.
A metric space approach to the information capacity of spike trains.

Metric spaces.

Metrics and spike trains.

- The framework for continuous version of information theory is a manifold, but perhaps that isn’t needed, perhaps it can be rephrased in terms of metric spaces.
The van Rossum metric.

- A spike train is a list of spike times.
  \[ u = \{u_1, u_2, \cdots, u_m\} \]

- Map spike trains to functions of \( t \)
  \[ u \mapsto f(t; u) = \sum_{i=1}^{m} h(t - u_i) \]

- \( h(t) \) is a kernel, here, it is a causal exponential function
  \[ h(t) = \begin{cases} 
  \exp(-t/\delta_T) & t > 0 \\
  0 & t \leq 0 
  \end{cases} \]

- Now
  \[ d(u, v) = \sqrt{\int dt [f(t; u) - f(t; v)]^2} \]
The van Rossum metric.

Two steps

- Maps from spike trains to functions using a filter.

- Use the metric on the space of functions.
A metric space approach to the information capacity of spike trains.

The van Rossum metric.
Comparing metrics

The basic idea is to use the candidate metric to cluster a set of spike trains, and to compare this clustering with a “gold standard”, namely, clustering the spike trains according to the stimuli that elicited them.

The scheme we use is a jack-knife calculation of a confusion matrix. The transmitted information \( \tilde{h} \) is used to score clustering with one, the highest, corresponding to perfect clustering.
How would information theory work on the metric space of spike trains?
Imagine . . .

- First let's ask how it would look if there were coordinates for spike trains.
- Imagine there is a space of spike trains with coordinates and all that.

- Imagine there is a coordinate for each length $L$ piece of spike train.
Imagine further . . .

- Imagine that each variable has independent additive Gaussian noise.

\[ X_i = Y_i + \eta \]
The $\chi$-distribution.

- The distance between two such vectors satisfies a $\chi$-distribution: $\mathbf{X} = (X_1, X_2, \ldots, X_k)$, $\mathbf{X}' = (X'_1, X'_2, \ldots, X'_k)$ has $|\mathbf{X} - \mathbf{X}'| \sim \chi(\sigma, k)$. 

\[ \chi(k = 2) \]

Gaußian

\[ \text{Gaußian} \]
A metric space approach to the information capacity of spike trains.

Noise on the metric space of spike trains.

The $\chi$-distribution.
A metric space approach to the information capacity of spike trains.

Noise on the metric space of spike trains.

Idea!

- Turn this around!²


Noise on the metric space of spike trains.

Idea!

- Turn this around!²
  - Propose this as the distribution of distances.
  - Calculate $k$ from the distribution and use this to work out $L$.

$$k = \frac{2\langle \zeta^2 \rangle^2}{\langle \zeta^4 \rangle - \langle \zeta^2 \rangle^2}.$$  

- Use the noise model to calculate information.

---

A metric space approach to the information capacity of spike trains.

Noise on the metric space of spike trains.

Idea!

- $k$ is a sort of noise dimension or effective dimension.
A metric space approach to the information capacity of spike trains.

Results.

$\chi$-distribution.

- Tested using the Anderson-Darling test.
A metric space approach to the information capacity of spike trains.

Results.

$k$ as a function of spike train length.

- $k$ should increase linearly with sample length.
Channel capacity.

The channel capacity for a single Gaussian variable $X$ is

$$C = \frac{1}{2} \log_2 \left(1 + \frac{\nu^2}{\sigma^2}\right) \text{ bits per time unit}$$

where $\sigma^2$ is the signal variance, usually taken to be the bound by the power constraint and $\nu^2$ is the noise variance.
Information theory - this works.

- Model the spike train as a Gaussian channel but re-express the calculations in terms of distance based quantities!
Results.

Information theory - this works.

For example,

- If $X$ and $X'$ are iid Gaussian variables with variance $\sigma^2$ their difference is Gaussian with variance $\sigma_d^2 = 2\sigma^2$. 
A metric space approach to the information capacity of spike trains.

Results.

Channel capacity.

\[ C = \frac{1}{2} \log_2 \left( \frac{\xi_d^2}{\sigma_d^2} \right) \text{ bits per } L. \]

where \( \xi_d^2 \) is the signal variance and \( \sigma_d^2 \) the noise variance and \( L = \text{(sample length)}/k \).
Variances.

- $\xi_d^2$ and $\sigma_d^2$ are calculated by least squares fit.
Channel capacity of the cells we looked at.
Information theory on the metric space.

- The noise model fits the data we have.
- Seems to be the natural arena for information theory calculations.
- The channel capacity theory is about encoding discrete information in a continuous signal.
  - What we actually need is distortion theory.
- A multi-neuron version is needed for populations.
- Most of all, need to apply to more data.
More general conclusions.

- Information theory - what’s the story with that?
- So, what is the space of spike trains?