Reverse-correlation method for receptive field estimation

Theoretical Neuroscience

Fabrizio Gabbiani
Division of Neuroscience
Baylor College of Medicine
One Baylor Plaza
Houston, TX 77030
e-mail:gabbiani@bcm.tmc.edu
1 Introduction

In the previous lectures, we described two methods to characterize the receptive field properties of visual neurons. The first method consists in presenting small spots of light or bars, depending on the shape of the receptive field, at various positions and recording the elicited response. The second method consists in presenting drifting sinusoidal gratings. The two methods are related for a linear system: the first method allows to compute directly the receptive field weighting function characterizing the transformation between an arbitrary stimulus and firing rate. The second method allows to compute the Fourier transform of the weighting function. We now introduce a third method called the reverse-correlation method based on presentation of random stimuli.

For simplicity we will only explain how the reverse-correlation method allows to determine the temporal receptive field of a neuron. The fundamental principle remains the same for spatio-temporal receptive fields.

2 Gaussian white noise stimuli

Stochastic processes. The reverse-correlation method is based on the use of stimuli that vary randomly in time. To take a concrete example, we will assume that the aim is to determine the temporal transfer function $w_t(\cdot)$ characterizing the center dynamics of an LGN neuron. The stimulus used to apply the reverse-correlation method would then be a time-varying random contrast sequence, $c(t)$, presented at the center of the receptive field. Just as the values taken by a random variable $x$ are part of an ensemble characterized by a probability distribution $p(x)$, such a random time-varying function is a concrete example of an ensemble of random functions that is characterized by the joint distribution $p(c_1, \ldots, c_n)$ of values that it takes at times $t_1, \ldots, t_n$. Such random functions are called stochastic processes. The mean value of a stochastic process, $\langle c(t) \rangle$, or the correlation function $\langle c(t_1)c(t_2) \rangle$ are defined just as for usual random variables. However, they now depend on one or two continuous temporal parameters ($t$ or $t_1$ and $t_2$, respectively).

Gaussian white noise. We now define a special stochastic process $c(t)$ that turns out to be central in the formulation of the reverse-correlation method. For this stochastic process, we assume that the average value of a random contrast fluctuation $c(t)$ is equal zero at any time point $t$: $\langle c(t) \rangle = 0$. Furthermore we choose the contrast values at different time points such that they are independent and identically distributed following a Gaussian distribution. The contrast values at two different time points $t_1$ and $t_2$ are uncorrelated and we assume that

$$\langle c(t_1)c(t_2) \rangle = \sigma^2 \delta(t_1 - t_2).$$  \hspace{1cm} (1)

Such a stimulus is called Gaussian white noise. It is an idealization that is convenient for theoretical arguments. In practice, it is approximated by discretizing time in steps
\( t_n = n \Delta t \) and by choosing independently for each time step a contrast value \( c(t_n) \) from a Gaussian distribution with zero mean and variance \( \sigma^2 / \Delta t \). This follows from eq. 1 and from the discrete approximation to the delta function:

\[
\delta(t_n) = \begin{cases} 
0 & \text{if } n \neq 0, \\
1/\Delta t & \text{if } n = 0.
\end{cases}
\]

An example of such a discrete random white noise contrast fluctuation is illustrated in Fig. 1. In the discrete approximation, each value of \( c(t_n) \) is chosen independently for each time step \( \Delta t \). This corresponds to a Nyquist frequency \( f_{Nyquist} = 1/2\Delta t \), and implies that contrast can change at any frequency between 0 and \( f_{Nyquist} \). Each frequency in this range is equally represented in a white noise stimulus. In the frequency domain, this corresponds to a flat (or white) power density, thus the name white noise.

**Stationarity.** An important property of Gaussian white noise and of a large class of stochastic processes is that their average statistical properties are independent of the time reference point chosen. Such stochastic processes are referred to as stationary. In the case of the mean value, this means that \( \langle c(t) \rangle = c_0 \) independent of \( t \). For the autocorrelation function this means that \( \langle c(t_1)c(t_2) \rangle \) depends only on the time difference \( \tau = t_2 - t_1 \). The autocorrelation function is therefore usually defined as

\[
R_{cc}(\tau) = \langle c(t)c(t + \tau) \rangle.
\]

**Ergodicity.** Another important property of Gaussian white noise is that averages over the ensemble of contrast stimuli at a fixed time point can be replaced by an average over time on any contrast sample function \( c_0(t) \) from the Gaussian white noise stochastic ensemble. For example, in the case of the autocorrelation function \( R_{cc}(\tau) \):

\[
R_{cc}(\tau) = \langle c(t)c(t + \tau) \rangle = \frac{1}{T} \int_0^T dt c_0(t)c_0(t + \tau).
\]

This means that each sample function is "sufficiently diverse" to be representative of the variability that is encountered across the whole sample of contrast functions belonging to the stochastic ensemble. In the case of neurophysiological experiments replacing ensemble averages by time averages is convenient since it allows to use a single stimulus presentation instead of repeated presentations of stimuli from the same ensemble. Ergodicity can only hold in the case of stationary (time-invariant) signals.

### 3 Reverse-correlation for linear encoding systems

An important application of Gaussian white noise stimuli is that they allow to determine the receptive field weighting function of a neuron in a fast and efficient way. To see how
this arises, let us first define the firing rate modulation due to the contrast stimulus $c(t)$ as,

$$f(t) = \langle x(t) - f_{\text{mean}} \rangle_c,$$  \hfill (2)

where $x(t) = \sum_{i=1}^{N} \delta(t - t_i)$ represents the spike train response to a single presentation of the contrast waveform $c(t)$. The average $\langle \cdot \rangle_c$ is taken over repeated presentations of the white noise contrast waveform $c(t)$. The number $f_{\text{mean}}$ is the mean firing rate of the cell.

Let us now assume as in the previous lectures that contrast changes are encoded linearly by firing rate changes. In terms of eq. 2 above this means that,

$$f(t) = \langle x(t) - f_{\text{mean}} \rangle_c = \int dt_0 w_i(t - t_0) c(t_0).$$

If we multiply both sides of this equation by the stimulus $c(t)$ and average over the different random contrast waveform of the stochastic ensemble, we immediately see that the right hand side is proportional to $w_i$:

$$\int dt_0 w_i(t + \tau - t_0) \langle c(t)c(t_0) \rangle = \int dt_0 w_i(t + \tau - t_0) \sigma^2 \delta(t - t_0) = \sigma^2 w_i(\tau).$$

To obtain from the left hand side an expression that depends explicitly on the time of spike occurrences, we first observe that

$$\langle (x(t + \tau) - f_{\text{mean}}) c(t) \rangle = \langle x(t + \tau) c(t) \rangle,$$

because contrast changes average to zero: $\langle c(t) \rangle = 0$. We can now replace the average over the Gaussian white noise contrast ensemble by a time average assuming that the ergodicity property holds:

$$\langle x(t + \tau) c(t) \rangle = \frac{1}{T} \int_0^T dt \ x(t + \tau)c(t).$$

If we now plug in this equation the delta function spike train, we see that

$$\frac{1}{T} \int_0^T dt \ x(t + \tau)c(t) = \frac{1}{T} \int_0^T dt \sum_{i=1}^{N} \delta(t + \tau - t_i)c(t)$$

$$= \frac{1}{T} \sum_{i=1}^{N} c(t_i - \tau)$$

$$= f_{\text{mean}} \frac{1}{N} \sum_{i=1}^{N} c(t_i - \tau).$$

Summing up,

$$w_i(\tau) = \frac{f_{\text{mean}}}{\sigma^2} \left( \frac{1}{N} \sum_{i=1}^{N} c(t_i - \tau) \right).$$ \hfill (3)
In other words, the transfer function \( w_t \) at time \( \tau \) is given by the spike-triggered average of the contrast values \( \tau \) msec prior to each spike. Because of the time reversal between the left and right hand side of eq. (3), this method is called reverse-correlation (Fig. 2).

**Assumptions and equivalence of reverse-correlation with other receptive field mapping techniques.** Several assumptions on neuronal responses are implicit when reverse-correlation is applied to determine the receptive field properties of a neuron. Besides the assumptions of ergodicity and stationarity mentioned above, another important assumption is that the response properties of a neuron do not depend on the frequency content of the stimulus. If this were not the case, the receptive fields mapped by presenting repeatedly sinusoidal stimuli at various frequencies could turn out to be very different from those obtained with a stimulus set like white noise in which all frequencies are presented simultaneously. In practice, the temporal receptive fields obtained with these different methods have yielded comparable results (Fig. 3). One advantage of using white noise is its simplicity: only a single stimulus type is used throughout the experiment. The frequency content (\( f_{Nyquist} \)) of the white noise stimulus should however be chosen high enough to exceed the maximal frequency to which a neuron is sensitive. This does not pose substantial problems in the visual system where both spatial and temporal cut-off frequencies are reasonably low.

4 Extensions of the basic reverse-correlation technique

The reverse-correlation principle derived in the previous section can be extended to determine the spatial structure of a neuron’s receptive field in addition to its temporal structure. This is accomplished by presenting spatio-temporal white noise and by keeping track of where the stimulus was presented in space as well as in time. Although the reverse-correlation technique was first developed for the auditory system, this analysis technique has been by now been used successfully to estimate the weighting functions of many types of neurons, including retinal, LGN and cortical simple cell receptive fields, for example. An example of a simple cell receptive field obtained by reverse-correlation is illustrated in Fig. 3.

**Reverse-correlation and static non-linearities.** An interesting property of reverse-correlation is that it can be applied to obtain the linear weighting function of a receptive field even in cases where there is a strong rectification of the response or, more generally, in the case where linear processing is followed by a static non-linearity such as those described in the lecture on “Encoding of stimuli by instantaneous firing rate”. This is for example the case in simple cells, where the firing rate is thought to be described by linear weighting followed by half-wave rectification.

**Reverse-correlation and motion-energy.** Other non-linear transformation such as the one required to model the response of complex cells as in the motion-energy model
will cause the basic reverse-correlation method to break down. The characterization of such receptive field properties through random stimuli is still possible, but requires the use of more complex random stimuli.

Figure Legends

**Figure 1.** Top: Sample of a white noise stimulus ($\sigma = 1$) sampled at $\Delta t = 0.5$ ms. Bottom: corresponding frequency power density.

**Figure 2.** Top: Sketch illustrating the implementation of the reverse correlation technique. Bottom: Equivalence of tuning curves of auditory fibers computed by reverse correlation and traditional threshold response measurements. The curve labeled RF was obtained by Fourier transforming the reverse correlation function obtained from an auditory nerve fiber in response to a white noise stimulus. The curve labeled TC is a standard threshold sensitivity curve measuring the sound intensity required at a given frequency to elicit a threshold response in the same auditory nerve fiber. Adapted from De Boer, J. of Dynamic Syst., Meas. and Control, 265-273, Sept. 1973.

**Figure 3.** Two dimensional spatial response profiles of two different simple cell receptive fields. A: This cell had a strong bright excitatory region and weaker dark inhibitory region with a somewhat longer extent. B: In this cell, the bright excitatory region is stronger than the dark inhibitory region but their sizes are comparable. Adapted from Jones and Palmer, J. Neurophysiol. 58 6:1187-1211, 1987.