INSTITUTE OF PHYSICS PUBLISHING NETWORK: COMPUTATION IN NEURAL SYSTEMS

Network: Comput. Neural Syst. **14** (2003) 673–702 PII: S0954-898X(03)65794-5

Measuring linear and quadratic contributions to neuronal response

Duane Q Nykamp[1]

Department of Mathematics, University of California, Los Angeles, CA 90095-1555, USA

E-mail: nykamp@math.umn.edu

Received 16 July 2002, accepted for publication 4 July 2003
Published 1 August 2003
Online at stacks.iop.org/Network/14/673

Abstract

We present a method to dissociate the sign-dependent (linear or odd-order) response from the sign-independent (quadratic or even-order) response of a neuron to sequences of random orthonormal stimulus elements. The method is based on a modification of the classical linear–nonlinear model of neural response. The analysis produces estimates of the stimulus features to which the neuron responds in a sign-dependent manner, the stimulus features to which the neuron responds in a sign-independent manner and the relative weight of the sign-independent response. We propose that this method could be used to characterize simple and complex cells in the primary visual cortex.

1. Introduction

A highly idealized model of neuronal response to a stimulus is the linear–nonlinear model. In this model, spiking probability is a linear function of the stimulus, composed with a sigmoidal nonlinearity to ensure nonnegative probabilities. The linear–nonlinear neuron behaves essentially like a linear system. It has virtually opposite responses to stimuli with opposite signs. Its response to a sum of stimuli can be largely predicted by its response to each stimulus individually.

Clearly most neurons, even in primary sensory regions, are not well represented by a linear–nonlinear model. In the primary visual cortex, for example, only simple cells respond similarly to a linear–nonlinear model. Complex cell response is more fundamentally nonlinear and cannot be captured by a linear–nonlinear model.

One feature of complex cells is their indifference to the contrast sign of the visual stimulus. For example, an idealized complex cell responds similarly to a black or a white bar on a grey background. We extend the linear–nonlinear framework to capture this sign-independent response. We allow the neuron's response to be a linear function not only of the stimulus values

[1] Present address: School of Mathematics, University of Minnesota, 206 Church Street, Minneapolis, MN 55455, USA.

0954-898X/03/040673+30$30.00

but also of locally squared values (i.e. pixel-by-pixel squaring) of the stimulus (composed with a sigmoidal nonlinearity as before). This squaring gives sign independence with only a small modification of the linear–nonlinear model.

We view this linear–quadratic–nonlinear model strictly as a phenomenological model of the sign independence and do not view a neural basis for squaring as essential for the validity of the approach. In essence, we use the linear–quadratic formalism simply to separate influences that are odd-order in the input from influences that are even-order. The advantage of this model is that it can be completely reconstructed from its response to a random sequence of orthonormal stimulus elements (cf [20]), similar to the reconstruction of the linear–nonlinear model from a white noise stimulus (e.g. [3, 16]).

The reconstruction process leads to estimates of the stimulus features to which the neuron responds in a sign-dependent and sign-independent manner. In addition, we obtain a measure, called the quadratic index of nonlinearity (QIN), that indicates the fraction of neural response that is independent of contrast sign (i.e. even-order in the input). The QIN is a quantitative gauge of nonlinear receptive field organization that can be estimated from neural response to a large class of stimuli (random sequences of orthonormal stimulus elements).

In section 2, we specify the proposed linear–quadratic–nonlinear model and define the random orthonormal stimulus. In section 3, we analyse the model to determine a method to reconstruct its parameters from measured spike times. In section 4, we demonstrate the ability of the method to determine the model even when the assumptions underlying the analysis are relaxed. We discuss the results in section 5.

2. Specification of model

2.1. Model of neural response

A traditional model for simple cells in the primary visual cortex is a linear function of the input composed with a static nonlinearity. If we let the response R of a neuron be 1 if the neuron spikes at a given time point and be 0 otherwise, then this model can be written as[2]

$$\Pr(R = 1|\boldsymbol{X} = \boldsymbol{x}) = g(\boldsymbol{h} \cdot \boldsymbol{x}), \tag{1}$$

where the vector $\boldsymbol{X}$ represents the recent spatio-temporal input, such as the pixel values for each refresh of a computer monitor. The linear function of the input is represented by the convolution with a kernel $\boldsymbol{h}$. By writing the convolution as a dot product $\boldsymbol{h} \cdot \boldsymbol{X}$, we are implicitly viewing the temporal index of the input as going backward in time. The function $g(\cdot)$ is some given nonlinear function.

Note that $\boldsymbol{X}$ does not represent the entire input sequence over the course of an experiment. It represents only the *recent* input so that the dot product $\boldsymbol{h} \cdot \boldsymbol{X}$ includes the entire memory of the system.

We propose an extension that accounts for the sign-independent response of complex cells. We allow the argument of the nonlinearity to include a linear function of locally squared values of the input, i.e. pixel-by-pixel squared input values. We add a term to model (1), obtaining the spiking probability

$$\Pr(R = 1|\boldsymbol{X} = \boldsymbol{x}) = g\left(\sqrt{1-\alpha}\boldsymbol{h}_1 \cdot \boldsymbol{x} + \sqrt{\alpha}\boldsymbol{h}_2 \cdot \boldsymbol{x}^2\right). \tag{2}$$

For local squaring, we use the notation $\boldsymbol{x}^2$ to indicate the vector with components that are the components of $\boldsymbol{x}$ squared. The kernels $\boldsymbol{h}_1$ and $\boldsymbol{h}_2$ are normalized so that the variances of $\boldsymbol{h}_1 \cdot \boldsymbol{X}$ and $\boldsymbol{h}_2 \cdot \boldsymbol{X}^2$ are both 1.

[2] We use capital letters to indicate random quantities.

The parameter α, with $0 \leqslant \alpha \leqslant 1$, is the QIN. It measures a particular departure from linearity of the neural response. When $\alpha = 0$, we regain the traditional linear–nonlinear model (equation (1)) of a simple cell. When $\alpha = 1$, the neuron responds to the stimulus independent of its sign, like a prototypical complex cell. With intermediate values of the QIN, the neural response contains both sign-dependent and sign-independent elements. When $\alpha = 1/2$, the variance of the sign-dependent and sign-independent components of the nonlinearity argument are equal. The parameter α is introduced into the equation so that the variance of the nonlinearity argument is always 1 (shown below).

We demonstrate below that model (2), including the QIN, can be reconstructed from the spike response to a random sequence of orthonormal stimulus elements. Thus, the QIN can be easily estimated from experiments to index the nonlinearity of the neurons' response. As shown below, the value of the QIN is not limited to neurons whose response is well described by equation (2) because the final equation for α can be interpreted in a model-independent manner.

2.2. *Description of stimulus*

The motivation behind this analysis is the random grating stimulus developed by Ringach *et al* [20]. In their experiments, sinusoidal gratings specified by a set of Hartley basis functions are randomly flashed on a computer monitor. We exploit the fact that the Hartley basis functions are orthonormal to reconstruct our linear–quadratic–nonlinear model (equation (2)) from such experiments.

In fact, orthonormality is the only condition we need from the Hartley basis functions. The analysis we describe applies to a random sequence from any set of stimulus elements as long as those elements are orthonormal. With such an orthonormal set, the required calculations can be performed completely in the space defined by the coefficients of the stimulus elements (the Fourier domain for the case when the stimulus set contained sinusoidal gratings). In that coefficient space, the stimulus is extremely simple, analogous to a one-dimensional random sequence of bright or dark dots. We will exploit this simple stimulus to reconstruct model (2) including the QIN.

We first describe the stimulus in general. The stimulus ensemble is composed of m orthonormal vectors $\{\hat{e}_j\}$ and their opposites $\{-\hat{e}_j\}$. By orthonormal, we mean that

$$\langle \hat{e}_j, \hat{e}_k \rangle = \begin{cases} 0 & \text{if } j \neq k, \\ 1 & \text{if } j = k, \end{cases} \tag{3}$$

where $\langle \cdot, \cdot \rangle$ is the standard vector inner product. We write the $2m$ vectors of the stimulus ensemble as

$$\{b\hat{e}_j : j = 1, \ldots, m \text{ and } b = -1, 1\},$$

where b is a binary variable denoting the sign of the stimulus element.

We require the stimulus ensemble to include members, called *blanks*, which are known to have no effect on the neuron's activity; the neuron's response is unaffected if a blank stimulus element were replaced by a completely blank stimulus or vice versa. Denote the set of indices corresponding to blanks by Λ_{blank} so that the set of blanks is $\{b\hat{e}_j : j \in \Lambda_{\text{blank}}, b = -1, 1\}$. For the actual stimulus, these members could be true blanks (vectors with all components zero), but the analysis is simpler if we treat these members as part of the orthonormal set.

The analysis is easiest to conceptualize if one views each member of the stimulus ensemble as a single dot in any of m positions along a line. We will proceed with our analysis as if this were the case. A general orthonormal stimulus ensemble is equivalent to this dot stimulus

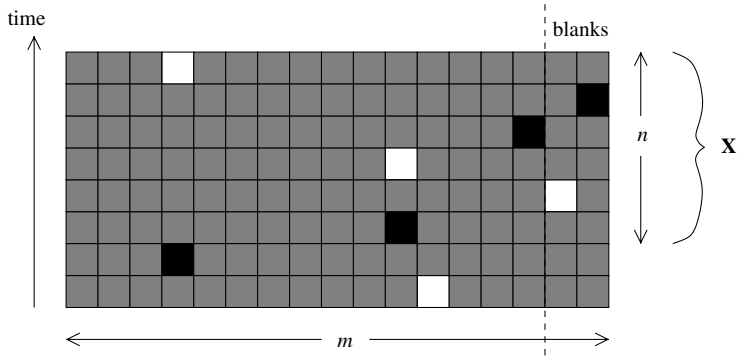

Figure 1. Example stimulus. Each horizontal line corresponds to a time point. Grey squares indicate 0 s, white squares indicates 1 s and black squares indicate −1 s. Squares to the right of the dotted line represent 'blanks' to which the cell does not respond. The stimulus vector $\boldsymbol{X}$ represents the stimulus at the previous n time steps.

when viewed in the space defined by the coefficients of the stimulus elements. We address the general case in more detail in section 3.6 and appendix A.2.

Let $\hat{e}_j$ be an m-dimensional vector that is zero except for the jth component so that it corresponds to a dot at position j. The stimulus member $b\hat{e}_j$ will be a white or black dot at position j, depending on the sign of b. It is known *a priori* that the neuron does not respond to dots in positions $j \in \Lambda_{\text{blank}}$. (Either these positions are well outside a neuron's receptive field or a completely blank stimulus was actually displayed instead of the dots.)

The input or stimulus to a neuron is a random sequence of vectors from the stimulus ensemble (see figure 1). Since the random stimulus and the neuronal response modelled below are stationary, we focus our attention at some particular time, which will be representative of any time. Assume that the response of a neuron at that time depends only on the previous n stimulus elements. Denote by $\boldsymbol{X}$ the mn-dimensional vector (really a matrix) of these stimulus elements. Note that $\boldsymbol{X}$ is only the recent stimulus corresponding to those n time points within a neuron's memory. In an actual experiment, one employs a much longer stimulus sequence, which we view as many realizations of the vector $\boldsymbol{X}$.

Given that each of the n previous time steps could contain one of $2m$ possible stimulus elements, the recent stimulus $\boldsymbol{X}$ could be $(2m)^n$ different combinations of stimulus elements. For each time step, the stimulus element is chosen randomly and independently, so that a given element is chosen with probability $1/(2m)$. Therefore, each possibility for $\boldsymbol{X}$ is equally likely, occurring with probability $1/(2m)^n$.

2.3. Enumeration of recent stimulus possibilities

To reconstruct model (2), we will compute expected values of the functions of the (recent) stimulus $\boldsymbol{X}$. To facilitate the computations, we use the following notation to enumerate the possible values of $\boldsymbol{X}$.

Denote by X^i_j the component of the input $\boldsymbol{X}$ at position j and time i (i.e. i time steps previous to the given time)[3]. X^i_j could be -1, 0 or 1, corresponding, for example, to a black, grey or white pixel, respectively.

Next, let $\boldsymbol{e}^i_j$ indicate the occurrence of a dot at position j and time i, i.e. $\boldsymbol{e}^i_j$ is a mn-dimensional vector that is zero except for the value 1 at position j and time i. A particular

[3] Throughout this paper, we will use subscripts to denote space indices and superscripts to denote time indices.

realization of $\boldsymbol{X}$ is a sequence of signed dots so that $\boldsymbol{X}$ can be written as a sum of vectors $\boldsymbol{e}^i_j$ with binary coefficients.

A particular realization of $\boldsymbol{X}$ can be achieved as follows. For a given time point i, roll a m-sided die to determine the position of the dot, then flip a coin to determine the sign of the dot. Repeat this procedure for each of the n time points. We record the sequence of dot locations in the vector $\boldsymbol{J}$ and the sequence dots signs in the vector $\boldsymbol{B}$. We denote the set of all possible sequences of dot positions by $\mathcal{J}^n_m$ and the set of all possible sequences of dot signs by $\mathcal{B}^n$.

Formally, $\boldsymbol{J}$ is an n-dimensional vector with each component $J^i \in \{1, 2, \ldots, m\}$ denoting the randomly chosen position at time i. To denote the set of all possible sequences of dot positions, $\mathcal{J}^n_m$ is the set all such n-dimensional vectors with components in $\{1, 2, \ldots, m\}$. Similarly, $\boldsymbol{B}$ is an n-dimensional vector with components $B^i \in \{-1, 1\}$ and $\mathcal{B}^n$ is the set of all such binary n-dimensional vectors. Note that $\mathcal{J}^n_m$ has m^n elements and $\mathcal{B}^n$ has 2^n elements.

For a given sequence of dot positions $\boldsymbol{J}$ and dot signs $\boldsymbol{B}$, the recent input $\boldsymbol{X}$ can be written as

$$\boldsymbol{X} = \sum_{i=1}^{n} B^i \boldsymbol{e}^i_{J^i},$$

as we simply add up dots at time i with position J^i and sign B^i. We write the $(2m)^n$ equally likely possibilities for $\boldsymbol{X}$ as

$$\boldsymbol{X} \in \left\{ \sum_{i=1}^{n} B^i \boldsymbol{e}^i_{J^i} : \boldsymbol{J} \in \mathcal{J}^n_m \& \boldsymbol{B} \in \mathcal{B}^n \right\}. \tag{4}$$

3. Analysis of model

3.1. Overcoming a singular stimulus auto-covariance

For the linear–nonlinear model of equation (1), cross-correlation methods are commonly used to reconstruct the model [1, 3, 5–7, 9, 12, 17, 20, 24]. In the special case where $g(\cdot)$ is a linear function, these techniques can be used to reconstruct the model even when the stimulus is as complicated as natural stimuli [19, 23].

To attempt to use these methods for our model and stimulus, one can reduce the linear–quadratic–nonlinear model of equation (2) to a linear–nonlinear model of equation (1) through a mathematical trick. One simply views the stimulus as including not only $\boldsymbol{X}$ but also the squared values $\boldsymbol{X}^2$. The model of equation (2) is a linear–nonlinear model of this enlarged 'virtual' stimulus. The statistics of this enlarged stimulus follow directly from the statistics of the physical stimulus $\boldsymbol{X}$.

Unfortunately, a direct application of the existing methods to this system fails. First, since the stimulus is not rotationally invariant, one must worry that a nonlinear $g(\cdot)$ could skew the estimation of the linear kernel [3] (but cf [20]). Second, and more importantly, the solution is ill-posed due to the statistical structure of the random orthogonal stimulus sequence.

The ill-posedness results from the strong dependence among pixel values in a given frame (viewing the stimulus as a sequence of dots, where each dot corresponds to one pixel, as in figure 1). If, for a given frame (horizontal line in figure 1), one observes a nonzero pixel at a given location, one immediately knows that all other pixels are zero (grey) because only one pixel is 'on' per frame. In the case where model (2) is linear in the physical stimulus ($\alpha = 0$), this dependence is benign because of cancellation between positive and negative pixel values [20]. For nonzero α, the quadratic term removes sign information so that this dependence among pixels becomes critical.

To illustrate the problem, we examine the special case when $\alpha = 1$. In this case, the system is a linear–nonlinear function of the squared pixel values

$$\Pr(R = 1|\boldsymbol{X} = \boldsymbol{x}) = g(\boldsymbol{h}_2 \cdot \boldsymbol{x}^2),$$

which we write as

$$\Pr(R = 1|\boldsymbol{Z} = \boldsymbol{z}) = g(\boldsymbol{h}_2 \cdot \boldsymbol{z}), \tag{5}$$

where $\boldsymbol{Z} = \boldsymbol{X}^2$. In this notation where $\boldsymbol{Z}$ appears as the stimulus, we are viewing all nonzero pixels as being positive, i.e. turning all black pixels of figure 1 into white pixels.

Ignoring the fact that this stimulus is not rotationally invariant (so that a nonlinear $g(\cdot)$ may skew this estimation procedure [3]), we could attempt to recover the kernel $\boldsymbol{h}_2$ from stimulus–spike correlations. To accomplish this, we must multiply by the inverse of the stimulus auto-covariance matrix. However, because of the strong dependence among pixel values of this stimulus, the auto-covariance matrix for $\boldsymbol{Z}$ is singular. (One does not encounter this problem when $\alpha = 0$ because oppositely signed values of $\boldsymbol{X}$ cancel each other, leaving an auto-covariance matrix for $\boldsymbol{X}$ that is proportional to the identity.)

In the special case where the system's memory is equal to a single frame ($n = 1$), the auto-covariance matrix of $\boldsymbol{Z}$ is proportional to an $m \times m$ matrix with $m - 1$ on the diagonal and -1 everywhere else. (Recall that m is the number of dot locations.) Since a given row of the matrix is equal to -1 times the sum of all other rows, the matrix is singular. For the general case ($n > 1$), since frames are independent, the auto-covariance matrix is block-diagonal, where each block is the above singular matrix.

The singularity of the stimulus auto-covariance matrix is not limited to the case when $\alpha = 1$. To apply the standard methods to model (2) with general α, one must use the enlarged 'virtual' stimulus mentioned above that includes both $\boldsymbol{X}$ and $\boldsymbol{Z}$. As $\boldsymbol{X}$ and $\boldsymbol{Z}$ are uncorrelated, the enlarged stimulus auto-covariance matrix contains the same rows as those for $\boldsymbol{Z}$ (padded with zeros). Hence, the auto-covariance matrix is still singular.

This singularity was recognized by Ringach *et al* [18] when they attempted a reverse-correlation procedure nearly identical to correlating with $\boldsymbol{Z}$. (They ignored the phase information of Hartley basis functions which includes ignoring the sign information of $\boldsymbol{X}$.) To compensate, they introduced the notion of blanks, defined exactly as our blank stimulus elements. By using the response to blanks as a baseline, they were able to circumvent the singularity in the auto-covariance matrix.

The manner in which to compensate for the strong dependence among stimulus pixels depends on the form of the nonlinearity $g(\cdot)$. Ringach *et al* [18] subtracted off the logarithm of the response to blanks, justifying the logarithm based on the detection of deviations from the baseline. Their use of the logarithm is essentially equivalent to assuming an exponential form of $g(\cdot)$, which motivated our use of the exponential function for $g(\cdot)$ (see below).

Besides the above justification, the crucial reason for using an exponential $g(\cdot)$ is practical. An exponential form for $g(\cdot)$ allows us to solve a large system of nonlinear equations explicitly, both avoiding the challenge of a numerical solution and providing an intuitive final solution. Moreover, our simulations, below, demonstrate that the result is robust to changes in the shape of the nonlinearity $g(\cdot)$.

3.2. Outline of procedure

The procedure we implement is similar to our approach to reconstructing a linear–nonlinear model in response to white noise [16]. Although the structure of the stimulus and the quadratic term of model (2) require more detailed manipulations, we again simply calculate analytic expressions for experimentally measurable stimulus–spike statistics and solve those equations for the model parameters.

We obtain expressions for the mean firing rate $E\{R\}$, the correlation of the spikes with the positive stimuli (e.g. white dots of figure 1), and the correlation of the spikes with the negative stimuli (e.g. black dots). Recall that each stimulus component X_j^i, viewed as the coefficient of the jth stimulus element, can assume the values -1, 0 and 1. The expression $X_j^i(X_j^i + 1)$ is 2 when $X_j^i = 1$ and is zero otherwise. The expression $X_j^i(X_j^i - 1)$ is 2 when $X_j^i = -1$ and is zero otherwise. As a result, $E\{X_j^i(X_j^i + 1)R\}$ is the correlation of the spikes with the positive stimuli and $E\{X_j^i(X_j^i - 1)R\}$ is the correlation of the spikes with the negative stimuli[4].

To isolate the sign-dependent or linear portion of the response, we subtract the correlation with the negative stimuli from the correlation with the positive stimuli, compensating for an exponential nonlinearity. This subtraction eliminates the sign-independent response. The result is proportional to the linear kernel $\boldsymbol{h}_1$ (see equation (2)).

Then, to isolate the sign-independent, or even-order, portion of the response, we add the correlations with both positive and negative stimuli, compensating for an exponential nonlinearity. After accounting for the baseline response to blanks, we will have eliminated the linear response. The result is proportional to the quadratic kernel $\boldsymbol{h}_2$.

We calculate the QIN α as the fraction of variance contained in the sign-independent portion. By factoring in the mean firing rate $E\{R\}$, we can also estimate parameters describing the exponential function $g(\cdot)$.

3.3. Calculation of stimulus–spike statistics

To carry out this procedure, we calculate analytic expressions for the mean firing rate and the correlation of spikes with positive and negative stimuli. The calculations are based on model (2) and the statistics of the recent stimulus $\boldsymbol{X}$. We also derive equations resulting from our normalization condition that $\boldsymbol{h}_1 \cdot \boldsymbol{X}$ and $\boldsymbol{h}_2 \cdot \boldsymbol{X}^2$ have unit variance. The goal of this analysis is to solve for the model parameters from measurements of these stimulus–spike statistics.

3.3.1. The mean firing rate. We first calculate the expression for the mean firing rate $E\{R\}$. Given model (2), the mean firing rate is simply the average of the spiking probability

$$g\left(\sqrt{1-\alpha}\boldsymbol{h}_1 \cdot \boldsymbol{X} + \sqrt{\alpha}\boldsymbol{h}_2 \cdot \boldsymbol{X}^2\right)$$

taken over all $(2m)^n$ possibilities of $\boldsymbol{X}$. As described in section 2.3, a particular realization of the recent input $\boldsymbol{X}$ is given by a sequence of dot positions $\boldsymbol{J}$ and dot signs $\boldsymbol{B}$ and can be written as

$$\boldsymbol{X} = \sum_{i=1}^{n} B^i \boldsymbol{e}^i_{J^i}, \tag{6}$$

where $\boldsymbol{e}_j^i$ is the vector denoting a dot at position j at time i. The component X_j^k is nonzero only if $j = J^k$, in which case $X_j^k = B^k$.

Consequently, the convolution $\boldsymbol{h}_1 \cdot \boldsymbol{X}$ (written as a dot product due to our backward time convention for $\boldsymbol{X}$) is simply

$$\boldsymbol{h}_1 \cdot \boldsymbol{X} = \sum_{k=1}^{n}\sum_{j=1}^{m} h_{1,j}^k X_j^k = \sum_{k=1}^{n} B^k h_{1,J^k}^k.$$

[4] We use the term correlation loosely, as we have not even subtracted off the product of the expected values.

(We denote by $h^k_{1,j}$ the component of $\boldsymbol{h}_1$ corresponding to time point k and space point j.) Since $(X^k_{J^k})^2 = (B^k)^2 = 1$, the convolution $\boldsymbol{h}_2 \cdot \boldsymbol{X}^2$ is

$$\boldsymbol{h}_2 \cdot \boldsymbol{X}^2 = \sum_{k=1}^{n} \sum_{j=1}^{m} h^k_{2,j} (X^k_j)^2 = \sum_{k=1}^{n} h^k_{2,J^k}.$$

All $(2m)^n$ possible values of $\boldsymbol{X}$ are given by the set of all possible sequences of dot positions ($\boldsymbol{J} \in \mathcal{J}^n_m$) combined with the set of all possible sequences of dot signs ($\boldsymbol{B} \in \mathcal{B}^n$). Taking the average of the spiking probabilities over these $(2m)^n$ values, we find that the mean firing rate is

$$\begin{aligned} E\{R\} &= E\left\{ g\left(\sqrt{1-\alpha}\boldsymbol{h}_1 \cdot \boldsymbol{X} + \sqrt{\alpha}\boldsymbol{h}_2 \cdot \boldsymbol{X}^2 \right) \right\} \\ &= \frac{1}{(2m)^n} \sum_{\boldsymbol{J} \in \mathcal{J}^n_m} \sum_{\boldsymbol{B} \in \mathcal{B}^n} g\left(\sqrt{1-\alpha} \sum_{k=1}^{n} B^k h^k_{1,J^k} + \sqrt{\alpha} \sum_{k=1}^{n} h^k_{2,J^k} \right). \end{aligned} \tag{7}$$

The sum defining the mean response is intractable for any practical use. Even with just modestly sized m and n, the number of terms in the sum is enormous. For example, with $m = 100$ and $n = 5$, the number of terms is $(2m)^n = 3.2 \times 10^{11}$. We must approximate the sum to make it manageable.

To calculate $E\{R\}$, we assume that

$$Y_R = \sqrt{1-\alpha}\boldsymbol{h}_1 \cdot \boldsymbol{X} + \sqrt{\alpha}\boldsymbol{h}_2 \cdot \boldsymbol{X}^2 \tag{8}$$

can be approximated by a Gaussian random variable. We calculate its mean μ_R and standard deviation σ_R and approximate

$$E\{R\} = E\{g(Y_R)\} \approx \frac{1}{\sigma_R \sqrt{2\pi}} \int g(y) \exp\left(-\frac{(y-\mu_R)^2}{2\sigma_R^2} \right) \mathrm{d}y.$$

In general, we would not expect n to be large enough for the central limit theorem to justify our approximation. The large sum in equation (7) merely means that partial sums approximating $E\{R\}$ should be nearly Gaussian. Nonetheless, approximating Y_R as Gaussian yields good results even in cases where this approximation is not justified, as demonstrated in section 4.2.

We divide Y_R into the component from $\boldsymbol{h}_1$ and the component from $\boldsymbol{h}_2$:

$$Y_R = \sqrt{1-\alpha} Y_{R1} + \sqrt{\alpha} Y_{R2}, \tag{9}$$

where

$$Y_{R1} = \boldsymbol{h}_1 \cdot \boldsymbol{X} \qquad \text{and} \qquad Y_{R2} = \boldsymbol{h}_2 \cdot \boldsymbol{X}^2. \tag{10}$$

As stated in the definition of model (2), the kernels $\boldsymbol{h}_1$ and $\boldsymbol{h}_2$ are normalized so that the variance σ^2_{R1} of Y_{R1} and the variance σ^2_{R2} of Y_{R2} are both 1. Since Y_{R1} and Y_{R2} are uncorrelated (though clearly not independent, see appendix A.1), the variance of Y_R is the weighted sum of their variances

$$\sigma^2_R = (1-\alpha)\sigma^2_{R1} + \alpha\sigma^2_{R2} = 1.$$

The definition of α was motivated by this result, ensuring that, independent of α, the variance of the nonlinearity argument Y_R is always one.

As shown in appendix A.1, the mean μ_{R1} of Y_{R1} is zero and the mean μ_{R2} of Y_{R2} is

$$\mu_{R2} = \sum_{k=1}^{n} \bar{h}^k_2, \tag{11}$$

where $\bar{h}_2^k$ is the average value of $\boldsymbol{h}_2$ at time point k

$$\bar{h}_2^k = \frac{1}{m}\sum_{j=1}^{m} h_{2,j}^k. \tag{12}$$

Hence, the mean μ_R of Y_R is simply the sum

$$\begin{aligned}\mu_R &= \sqrt{1-\alpha}\,\mu_{R1} + \sqrt{\alpha}\,\mu_{R2} \\ &= \sqrt{\alpha}\sum_{k=1}^{n} \bar{h}_2^k.\end{aligned} \tag{13}$$

When we approximate Y_R as a Gaussian random variable with mean μ_R and variance $\sigma_R^2 = 1$, the mean firing rate of our model neuron is

$$E\{R\} = \frac{1}{\sqrt{2\pi}}\int g(y)\mathrm{e}^{-(y-\mu_R)^2/2}\,\mathrm{d}y. \tag{14}$$

3.3.2. Stimulus–spike correlations. We repeat the same procedure for the correlation of spikes with the positive stimuli at position j and time i. We obtain an expression for $E\{X_j^i(X_j^i+1)R\}$ as an average of

$$X_j^i(X_j^i+1)g\left(\sqrt{1-\alpha}\boldsymbol{h}_1\cdot\boldsymbol{X} + \sqrt{\alpha}\boldsymbol{h}_2\cdot\boldsymbol{X}^2\right)$$

over all $(2m)^n$ possible realizations of the recent stimulus $\boldsymbol{X}$. The resulting sum includes only those stimulus realizations where a positive stimulus element occurred at position j and time i (i.e. when the vector $\boldsymbol{J}$ satisfies $J^i = j$ and the vector $\boldsymbol{B}$ satisfies $B^i = 1$). In those cases $X_j^i(X_j^i+1) = 2$; otherwise $X_j^i(X_j^i+1) = 0$. Hence

$$\begin{aligned}E\{X_j^i(X_j^i+1)R\} &= E\left\{X_j^i(X_j^i+1)g\left(\sqrt{1-\alpha}\boldsymbol{h}_1\cdot\boldsymbol{X} + \sqrt{\alpha}\boldsymbol{h}_2\cdot\boldsymbol{X}^2\right)\right\} \\ &= \frac{2}{(2m)^n}\sum_{\substack{J\in\mathcal{J}_m^n \\ J^i=j}}\sum_{\substack{\boldsymbol{B}\in\mathcal{B}^n \\ B^i=1}} g\left(\sqrt{1-\alpha}\sum_{k=1}^{n} B^k h_{1,J^k}^k + \sqrt{\alpha}\sum_{k=1}^{n} h_{2,J^k}^k\right) \\ &= \frac{1}{m}\frac{1}{(2m)^{n-1}}\sum_{\substack{J\in\mathcal{J}_m^n \\ J^i=j}}\sum_{\substack{\boldsymbol{B}\in\mathcal{B}^n \\ B^i=1}} g\left(\sqrt{1-\alpha}h_{1,j}^i + \sqrt{\alpha}h_{2,j}^i + Y_{X^iR}\right)\end{aligned} \tag{15}$$

where, for the last expression, we pulled out the known terms from time i. The remaining expression is

$$Y_{X^iR} = \sqrt{1-\alpha}\sum_{\substack{k=1 \\ k\neq i}}^{n} B^k h_{1,J^k}^k + \sqrt{\alpha}\sum_{\substack{k=1 \\ k\neq i}}^{n} h_{2,J^k}^k. \tag{16}$$

We obtain a similar expression for the correlation of the spikes with the negative stimuli, only with a negative sign in front of the $h_{1,j}^i$ term because $B^i = -1$. Thus, the stimulus–spike correlations are

$$E\{X_j^i(X_j^i \pm 1)R\} = \frac{1}{m}\frac{1}{(2m)^{n-1}}\sum_{\substack{J\in\mathcal{J}_m^n \\ J^i=j}}\sum_{\substack{\boldsymbol{B}\in\mathcal{B}^n \\ B^i=-1}} g\left(\pm\sqrt{1-\alpha}h_{1,j}^i + \sqrt{\alpha}h_{2,j}^i + Y_{X^iR}\right). \tag{17}$$

Note that the stimulus–spike correlations are written as $1/m$ times the average spiking probability over all $(2m)^{n-1}$ realizations of the recent stimulus, given that the stimulus for time point i is known.

As with the mean rate calculation, we approximate Y_{X^iR} (which is exactly Y_R with the ith time point removed) as Gaussian even though it is not strictly justified. The calculation of the mean μ_{X^iR} and variance $\sigma^2_{X^iR}$ of Y_{X^iR} are given in appendix A.1, although we end up canceling out these expressions in our final calculations below.

Approximating Y_{X^iR} as Gaussian, our two stimulus–spike correlation measures become

$$E\{X^i_j(X^i_j \pm 1)R\} = \frac{1}{m\sigma_{X^iR}\sqrt{2\pi}} \int g(y) \exp\left(-\frac{\left(y \mp \sqrt{1-\alpha}h^i_{1,j} - \sqrt{\alpha}h^i_{2,j} - \mu_{X^iR}\right)^2}{2\sigma^2_{X^iR}}\right) \mathrm{d}y. \tag{18}$$

Since $j = 1, \ldots, m$ and $i = 1, \ldots, n$, we have a total of $2mn$ equations of this form.

3.3.3. Normalization conditions for the kernels. The kernels $\boldsymbol{h}_1$ and $\boldsymbol{h}_2$ are normalized by the condition that the variances of Y_{R1} and Y_{R2} are 1. Appendix A.1 shows that the variance of Y_{R1} is

$$\sigma^2_{R1} = \frac{\|\boldsymbol{h}_1\|^2}{m} \tag{19}$$

and the variance of Y_{R2} is

$$\sigma^2_{R2} = \sigma^2_t(\boldsymbol{h}_2), \tag{20}$$

where $\|\boldsymbol{v}\|$ is the standard vector norm[5] of $\boldsymbol{v}$ and

$$\sigma_t(\boldsymbol{v}) = \left\{\sum_{k=1}^{n}\left[\frac{1}{m}\sum_{j=1}^{m}(v^k_j)^2 - \left(\frac{1}{m}\sum_{j=1}^{m}v^k_j\right)^2\right]\right\}^{1/2}. \tag{21}$$

The expression $\sigma^2_t(\boldsymbol{X})$ is the sum of the variance of each time slice of $\boldsymbol{v}$. For the proper normalization, we scale the kernels so that

$$\|\boldsymbol{h}_1\| = \sqrt{m} \qquad \text{and} \qquad \sigma_t(\boldsymbol{h}_2) = 1. \tag{22}$$

3.4. Model reconstruction

If we assume that the spiking probability of a neuron is given by equation (2), then equations (14) and (18) give the expected values of the mean rate and stimulus–spike correlations measured in response to a random sequence of orthonormal images. If we estimate these statistics by measuring the spike times of a neuron, we can calculate the model parameters that would lead to those measurements.

Equations (14), (18) and (22) give $2mn + 3$ conditions on the model parameters. From these conditions we need to estimate two $m \times n$ kernels ($\boldsymbol{h}_1$ and $\boldsymbol{h}_2$), the QIN α and the nonlinearity $g(\cdot)$. We need $2mn$ conditions to specify the $\boldsymbol{h}_j$. With one condition required for α, we have two conditions left to specify the nonlinearity $g(\cdot)$. We have enough information to pick $g(\cdot)$ out of a two-parameter family (just like the case of the linear–nonlinear model in response to white noise [16]).

3.4.1. Blanks enable well-posed problem. Given a particular choice for a two-parameter nonlinearity family of nonlinearities $g(\cdot)$, equations (14), (18) and (22) are a system of $2mn+3$ coupled nonlinear equations for $2mn + 3$ unknowns. One might imagine that, given estimates of the stimulus–spike statistics, they could be numerically solved for the model parameters.

[5] Given that we are using two indices, $\|\boldsymbol{v}\|^2 = \sum_{j,k}(v^k_j)^2$.

However, even if the system of equations were well-posed, solving the hundreds or thousands of coupled nonlinear equations could prove an exceedingly difficult task.

In fact, we run into the same problem discussed in section 3.1, where we attempted to solve the linear–nonlinear system using the expanded 'virtual' stimulus. Just as the stimulus auto-covariance matrix was singular, so also the above system of equations is ill-posed. If one assumes that $g(\cdot)$ is linear, the resulting matrix to be inverted is singular. From the $2mn + 3$ equations, there are really only $2mn + 3 - n$ linearly independent conditions (this case is equivalent to inverting the auto-covariance matrix of the expanded stimulus). The system cannot be unambiguously solved for the $2mn + 3$ model parameters.

For a nonlinear $g(\cdot)$, one could linearize around a solution to obtain a linear system valid close to the solution. The linear equations would be nearly identical to the equations for linear $g(\cdot)$, differing only because the slope $g'(\cdot)$ might vary for the different nonlinearity arguments. For a realistic nonlinearity, the slope $g'(\cdot)$ will change little across the limited range of nonlinearity arguments in our equations. If the matrix from this local problem is not singular, it will be nearly singular and the solution will be sensitive to small measurement errors.

We use the blank stimuli to add the necessary n conditions and make the solution for the model parameters well-posed. A position j simply corresponds to the position of a blank if $j \in \Lambda_{\text{blank}}$. Since we assume the neuron's response is not influenced by dots in a blank position, the correlation with positive blank stimuli is identical to the correlation with negative blank stimuli, $E\{X^i_j(X^i_j + 1)R\} = E\{X^i_j(X^i_j - 1)R\} = E\{(X^i_j)^2R\}$. For each time i, we average this stimulus–spike correlation over all blank locations $j \in \Lambda_{\text{blank}}$. (In cases where the blanks are not true blank stimuli, one might use a median rather than an average to remove the effect of the neuron responding to a few of the 'blank' stimuli, cf [18].)

The average stimulus–spike correlation for blank stimuli at time i is

$$E\{K^i\} = \frac{1}{|\Lambda_{\text{blank}}|} \sum_{j \in \Lambda_{\text{blank}}} E\{(X^i_j)^2 R\}, \tag{23}$$

where $|\Lambda_{\text{blank}}|$ is the number of positions corresponding to blank stimuli. The assumption that a neuron does not respond to a stimulus at position j is equivalent to assuming that $h^i_{1,j} = h^i_{2,j} = 0$ for all i. From equation (18) for zero kernels, the blank stimulus–spike correlation at time i is

$$E\{K^i\} = \frac{1}{m\sigma_{X^iR}\sqrt{2\pi}} \int g(y) \exp\left(-\frac{(y - \mu_{X^iR})^2}{2\sigma^2_{X^iR}}\right) \mathrm{d}y. \tag{24}$$

Equation (24) provides the n additional conditions needed to reconstruct the model parameters.

3.4.2. An exponential nonlinearity. Even with a well-posed problem, solving the approximately $2mn$ nonlinear equations numerically would be challenging. To avoid this difficulty, we choose a family of nonlinearities where the equations can be solved analytically. Since such a choice will yield explicit formulae for the model parameters, it will also give intuitive meaning to the final results.

Although a sigmoidal nonlinearity would, in general, be required to match the saturation of neural firing rates when driven strongly, the low firing rates of neurons in response to random stimuli makes modelling the saturation unnecessary. Instead, we use a simple exponential nonlinearity. Though this could lead to probabilities greater than one in equation (2), we assume the firing rate is sufficiently low and the temporal discretization sufficiently fine to preclude this exception.

3.4.3. Solution for model parameters. Assume that $g(y)$ is an exponential function

$$g(y) = \mathrm{e}^{\beta(y-\gamma)} \tag{25}$$

with parameters β and γ. By a completion of the square in the exponent, equations (18) and (24) become

$$E\{X_j^i(X_j^i \pm 1)R\} = \exp\Big(\beta\Big[\pm\sqrt{1-\alpha}h_{1,j}^i + \sqrt{\alpha}h_{2,j}^i\Big]\Big)E\{K^i\}$$

$$E\{K^i\} = \frac{1}{m}\exp\left(\frac{\beta^2\sigma_{X^iR}^2}{2} + \beta(\mu_{X^iR} - \gamma)\right)$$

so that

$$\beta\Big[\pm\sqrt{1-\alpha}h_{1,j}^i + \sqrt{\alpha}h_{2,j}^i\Big] = \log\left(\frac{E\{X_j^i(X_j^i \pm 1)R\}}{E\{K^i\}}\right). \tag{26}$$

The kernels are proportional to the sum or difference of the two signed forms of equation (26). Define $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ as the vectors formed by one-half the difference and sum, respectively:

$$v_{1,j}^i = \tfrac{1}{2}\log E\{X_j^i(X_j^i + 1)R\} - \tfrac{1}{2}\log E\{X_j^i(X_j^i - 1)R\}, \tag{27}$$

$$v_{2,j}^i = \tfrac{1}{2}\log E\{X_j^i(X_j^i + 1)R\} + \tfrac{1}{2}\log E\{X_j^i(X_j^i - 1)R\} - \log E\{K^i\}. \tag{28}$$

Note that the blanks dropped out of the equation for $\boldsymbol{v}_1$. The kernels $\boldsymbol{h}_1$ and $\boldsymbol{h}_2$ are proportional to the vectors $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$, respectively. Due to the normalization conditions from equation (22), the originally scaled kernels are

$$\boldsymbol{h}_1 = \frac{\sqrt{m}\boldsymbol{v}_1}{\|\boldsymbol{v}_1\|} \qquad \text{and} \qquad \boldsymbol{h}_2 = \frac{\boldsymbol{v}_2}{\sigma_t(\boldsymbol{v}_2)}, \tag{29}$$

where $\sigma_t(\boldsymbol{v})$ is defined by equation (21). The normalization conditions of $\boldsymbol{h}_1$ and $\boldsymbol{h}_2$ mean that

$$\|\boldsymbol{v}_1\|^2 = m\beta^2(1-\alpha) \qquad \text{and} \qquad \sigma_t^2(\boldsymbol{v}_2) = \beta^2\alpha$$

so that the QIN is given by

$$\alpha = \frac{m\sigma_t^2(\boldsymbol{v}_2)}{\|\boldsymbol{v}_1\|^2 + m\sigma_t^2(\boldsymbol{v}_2)}. \tag{30}$$

We also obtain for free that $\beta = [\|\boldsymbol{v}_1\|^2/m + \sigma_t^2(\boldsymbol{v}_2)]^{1/2}$. From a completion of the square in equation (14), the second nonlinearity parameter is $\gamma = \beta/2 + \mu_R - \log E\{R\}/\beta$, where μ_R is given by equation (13).

3.5. Model-independent interpretation

The value of the analysis is not limited to neurons whose response is well described by equation (2). Both the equations (29) for $\boldsymbol{h}_k$ and (30) for α can be interpreted apart from equation (2) used to derive the formulae.

Linear or odd-order behaviour implies responding to $X_j^i = 1$ in the opposite way as to $X_j^i = -1$ (though the nonlinearity eliminates true symmetry). Sign-independent or even-order behaviour implies responding to $X_j^i = 1$ in the same way as $X_j^i = -1$.

The formula (27) for $\boldsymbol{v}_1$ is simply the correlation with the positive stimuli minus the correlation with the negative stimuli, compensating for an exponential nonlinearity. The sign-independent response is removed by formula (27) so that $\boldsymbol{v}_1$ includes only the linear response.

The formula (28) for $\boldsymbol{v}_2$ is the sum of the correlations with both positive and negative stimuli, compensating for an exponential nonlinearity and the baseline response to blanks.

The linear response is removed by formula (28) so that $\boldsymbol{v}_2$ includes only the sign-independent response.

The linear and quadratic kernels are proportional to $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$, respectively. The normalization in equation (29) is unimportant except in the context of equation (2).

The QIN α defined by equation (30) is simply the relative weight of the sign-independent portion of the response to the total response. Although the motivation for weighting the linear portion $\boldsymbol{v}_1$ differently than the sign-independent portion $\boldsymbol{v}_2$ can be seen only from analysis of equation (2), the fundamental notion of α as the relative weight of a sign-independent response can be read directly from equation (30).

3.6. Further generalizations

3.6.1. General orthonormal stimuli. In the above analysis, we assumed that the orthonormal vectors $\hat{\boldsymbol{e}}_j$ composing the stimulus were simply dots. In appendix A.2, we show that the analysis carries through for any orthonormal set, such as Hartley functions (sinusoidal gratings) [20].

The only difference between dot stimuli and the general orthonormal stimuli is the calculation of the kernel $\boldsymbol{h}_2$. When the stimuli are not dots, we cannot calculate $\boldsymbol{h}_2$ in spatial coordinates. We can calculate $\boldsymbol{h}_2$ only in the coordinates given by the stimulus set, which we denote by $\tilde{\boldsymbol{h}}_2$. We can obtain the kernel $\boldsymbol{h}_1$ in stimulus coordinates (denoted by $\hat{\boldsymbol{h}}_1$) and in spatial coordinates as well as the QIN α. See appendix A.2 for more details.

3.6.2. Slower stimuli. In the model of equation (2), time must be discretized finely enough (intervals of the order of a millisecond) so that a neuron can spike at most once during a time interval. The analysis so far assumes that the stimulus frame changes every time step. In reality, experiments are designed with a much slower stimulus, with a frame interval in the tens of milliseconds. The slow stimulus is necessary not only due to limited monitor refresh rates but also to increase neural response to the random stimuli. Given a fixed maximal stimulus value, the effective power of the stimulus decreases with the frame rate, reducing neuronal response.

In appendix A.3, we generalize to the case when the stimulus is slower than the measurement interval. In the end, we must only account for the fact that we effectively over-count the time intervals by discretizing too finely. The kernels are the same up to an extra normalization factor that accounts for the over-counting, and the expression for the QIN α is unchanged.

4. Demonstration and test of method

To test the accuracy of the proposed method, we simulated neurons with spiking probability given by equation (2) and analysed the resulting spike times to reconstruct the model parameters. To facilitate the presentation of results, we used one-dimensional stimulus elements. We simulated neurons responding to random sequences of dots and to random sequences of sinusoids.

The kernels used in the simulation were chosen as one-dimensional caricatures of linear kernels measured in simple cells. They were given by

$$h_{k,j}(t) = t\mathrm{e}^{-t/\tau}\mathrm{e}^{-j^2/2\sigma^2}\cos(2\pi f j + \phi_k) \tag{31}$$

for $k = 1, 2$. The kernel and the stimulus contained 100 points in space. We discretized the kernel into 100 time slices at a resolution of 1 ms and normalized the results so that the $\boldsymbol{h}_k$

satisfied the appropriate normalization conditions (equations (22), (59) or (64)). Except where stated otherwise, we use the parameters $\tau = 20$ ms, $\sigma = 10$, $f = 0.04$, $\phi_1 = 0$ and $\phi_2 = \pi/2$.

In these demonstrations, we specify the nonlinearity parameters after μ_R (defined by equation (13)) has been subtracted from the nonlinearity argument. In other words, we specify the parameters of $\tilde{g}$ defined by

$$\tilde{g}(y) = g(y - \mu_R). \tag{32}$$

With this definition, the argument of $\tilde{g}$ always has zero mean and variance one, independent of $\boldsymbol{h}_k$ and α. By specifying $\tilde{g}$ instead of g, the firing rate of the neuron depends only on the nonlinearity and will not vary as we change $\boldsymbol{h}_k$ or α.

The above analysis is based on calculating the exact expected values of stimulus–spike statistics. In practice, one is limited to approximations of these expected values calculated from a finite data set. One must account for this limitation in order to achieve accurate results. First, the noisy estimates of the stimulus–spike statistics will lead to bias in the estimates of model parameters if one simply replaces the expected values with the noisy estimates. Second, one might encounter the situation where the estimate of $E\{X^i_j(X^i_j \pm 1)R\}$ is zero (i.e. no spikes occurred in any time bin i units of time after presentation of a white or black dot at position j). Given the assumption of an exponential nonlinearity, a zero spiking probability does not exist and equations (27) and (28) do not make sense when $E\{X^i_j(X^i_j \pm 1)R\} = 0$. Steps to reduce the bias from parameter estimates and correct zero correlation estimates are given in appendix B.

4.1. Initial demonstration

As an initial demonstration, we simulated a neuron with spiking probability given by equation (2) and standard kernel parameters specified above, using a temporal discretization of 1 ms. The nonlinearity was a exponential nonlinearity $\tilde{g}(y) = \mathrm{e}^{\beta(y-\gamma)}$ with $\beta = 1$ and $\gamma = 5$, so the average firing rate was approximately 11 spikes s^{-1}. The stimulus was a one-dimensional random dot stimulus with a period of 1 ms. Since the stimulus period equalled the temporal discretization, this example corresponds to the original derivation of the method.

We simulated the neuron with QIN α ranging between 0 and 1. We calculated the model parameters from the spike times after 10 min and 1 h of simulated time, obtaining approximately 7000 and 40 000 spikes, respectively. Figure 2 shows both the original simulated kernels and the kernels calculated with equations (29) for a few examples. With $\alpha = 0.5$, the calculated kernels match the qualitative features of the simulated kernels after 10 min (figures 2(A), (B)) and match quantitatively after 1 h (figures 2(C), (D)).

When $\alpha = 0.9$, most of the response is determined by the quadratic kernel $\boldsymbol{h}_2$ and little of the response is determined by the linear kernel $\boldsymbol{h}_1$. Consequently, after 10 min of simulated time, the calculated quadratic kernel matches the simulated kernel, but the calculated linear kernel is mostly noise (figures 2(E), (F)).

The method also accurately calculates the QIN using equation (30). The results from simulations with $\alpha = 0, 0.05, 0.1, \ldots, 1$ are summarized in figure 3. For the 10 min simulations (figure 3(A)), we also show results with different quadratic kernel phases $\phi_2 = 0, \pi/2, \pi$. Since the linear kernel phase is $\phi_1 = 0$, the results demonstrate the method works well, regardless of the relationship between the linear and quadratic kernels. Note that the QIN are relatively accurate after 10 min of simulated time even though the kernel estimates have not yet converged to the quantitatively correct results (cf figures 2(A), (B)).

4.2. Testing the Gaussian approximation

One might expect that the Gaussian approximation underlying the method would limit the applicability of the results. In the above example, the stimulus was fast compared to the

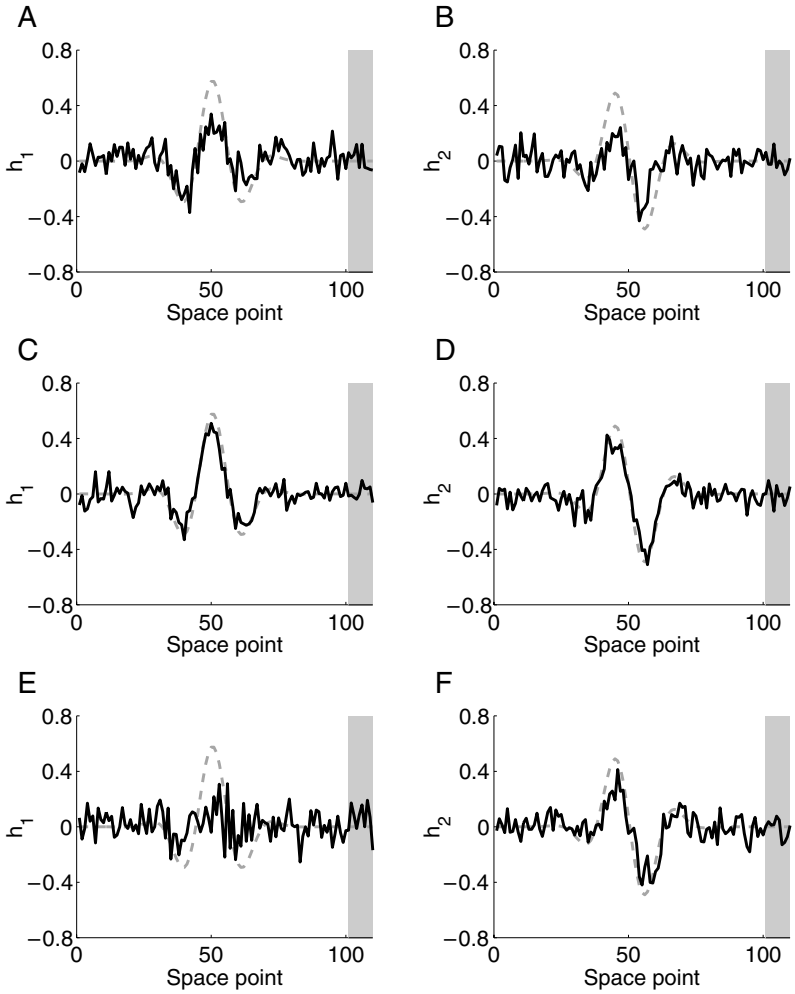

Figure 2. Snapshots of simulated (broken curve) and calculated (full curve) kernels from random dot stimulus. The shaded region indicates those points that represented blanks. (A)–(D) Linear and quadratic kernels with $\alpha = 0.5$ estimated from 10 min (A), (B) and 1 h (C), (D) of simulated time. (E), (F) Linear and quadratic kernels with $\alpha = 0.9$ estimated from 10 min of simulated time.

temporal width of the kernel so that the sum defining Y_R (equation (8)) contained many terms. In this case, the Gaussian approximation was justified.

To test the limitations of the method, we reduced the spatial and temporal scale of the kernels by setting $\tau = 2$ ms, $\sigma = 2$, and $f = 0.2$. In this case, the input to the nonlinearity Y_R contained few terms and most coefficients of these terms were nearly zero, invalidating any invocation of the central limit theorem. The results of these simulations are shown in figure 4. When $\alpha = 0.5$, the kurtosis, which measures the peakedness of a distribution, of Y_R was 6.5. (The kurtosis of a normal distribution is 0.) The distribution of Y_R, shown in figure 4(C), was clearly non-Gaussian, with a higher peak and longer tails. This breakdown of the Gaussian approximation increased the mean firing rate by nearly 30% to 14 spikes s^{-1}.

Nonetheless, the method reconstructed the qualitative features of the kernels after 10 simulated minutes (figures 4(A), (B)). The discrepancy between the simulated and calculated

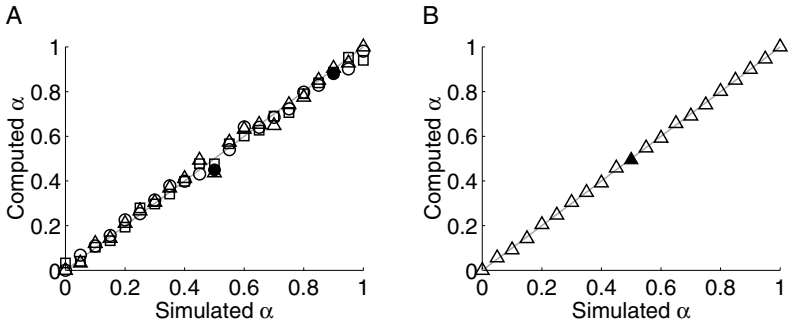

Figure 3. Comparison of simulated and computed QIN α from random dot stimulus. The line of equal α is indicated in grey. (A) Results from 10 min of simulated time. Circles indicate a quadratic kernel phase of $\phi_2 = \pi/2$, triangles indicate $\phi_2 = \pi$ and squares indicate $\phi_2 = 0$. (B) Results from 1 h of simulated time and $\phi_2 = \pi/2$. For both (A) and (B), filled symbols indicate simulations shown in figure 2.

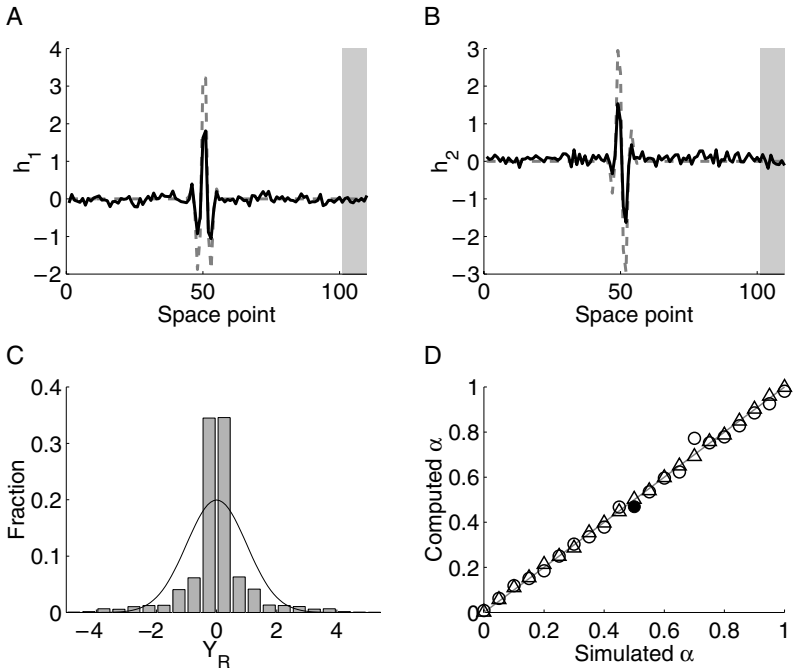

Figure 4. Results from simulations with narrow kernels. (A), (B) Panels are identical to figure 2(A), (B) except for the narrower kernels. Shown are kernels calculated from 10 min of simulated time with $\alpha = 0.5$. (C) The distribution of values of Y_R from the same simulation (grey bars). For comparison, a normal distribution of identical mean and variance is plotted with a black curve. Note that we subtracted off μ_R, the mean of Y_R, to be consistent with the modified nonlinearity $\tilde{g}(y)$. (D) Comparison of simulated and computed QIN α for simulations with narrow kernels. Results from 10 min (circles) and 1 h (triangles) of stimulated time are shown. The filled circle indicates the simulation of the previous panels.

kernels decreased but did not completely disappear after 1 simulated hour (not shown). The method also accurately estimated the QIN α over its entire range (figure 4(D)). Clearly, the Gaussian approximation required in the derivation is not necessary for accurate results.

4.3. Stimulus generalizations

To demonstrate the generalizations of section 3.6, we simulated the neuron's response to sequences of random sinusoidal gratings with a period of 10 ms. Since the temporal discretization was 1 ms, we used the generalization of the method to slower stimuli. The orthonormal elements $\hat{e}_k$ underlying the stimulus ensemble were one-dimensional Hartley functions normalized to unit length. If N is the number of space points, the Hartley functions are given by

$$\hat{e}_{k,j} = \frac{\cos(2\pi kj/N) + \sin(2\pi kj/N)}{\sqrt{N}}, \qquad \text{for } j = 0, \ldots, (N-1). \tag{33}$$

Using the kernels (equation (31)) with standard parameters, the neuron responds only to low spatial frequencies. In a one-dimensional analogue of the stimulus described in [20], we used only the 25 Hartley functions with $|k| \leqslant 12$. We also included four blank stimuli, which we viewed as members of the orthonormal set, denoting them by $\hat{e}_{13}, \ldots, \hat{e}_{16}$. Our stimulus ensemble was composed of these $m = 29$ orthonormal vectors and their opposites.

The results from these simulations are shown in figure 5. All the analysis is done in the space defined by the $\hat{e}_k$. Figures 5(A), (B) shows the resulting linear and quadratic kernels after 10 min of simulated time and $\alpha = 0.5$. Ignoring blanks, the linear kernel is transformed into spatial coordinates and shown in figure 5(C). Note that the estimate of the linear kernel is much less noisy than the estimate in figure 2(A). This improvement is due both to the smaller stimulus ensemble and the slower stimulus[6].

4.4. Other nonlinearities

We tested the robustness of the method to deviations in the shape of the nonlinearity. We simulated the response of a neuron with a power law nonlinearity

$$\tilde{g}(y) = \begin{cases} Ay^b & \text{if } y > 0 \\ 0 & \text{otherwise} \end{cases}$$

and with a error function nonlinearity

$$\tilde{g}(y) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{y - T}{\varepsilon\sqrt{2}}\right)\right]$$

where $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \mathrm{e}^{-t^2}\,\mathrm{d}t$. For the power law we used parameters $A = 0.02$ and $b = 2$, yielding an average firing rate of 10 spikes s^{-1}. For the error function, we used $T = 3.2$ and $\varepsilon = 1$, yielding an average firing rate of 13 spikes s^{-1}. We simulated the neurons in response to random dots with a period of 10 ms.

Since the nonlinearities were not exponentials, we observed slight discrepancies in the calculated values of the kernels, primarily with the quadratic kernel. The power law nonlinearity caused the largest deviations. An example with $\alpha = 0.5$ and 1 h of simulated time is shown in figures 6(A), (B). Since the noise is low due to the length of the simulation, the undershoot at the peaks of the quadratic kernel is clear. In all cases, the method accurately reproduced the qualitative features of the kernels.

The method also accurately estimated the QIN α as shown in figures 6(C) (power law) and (D) (error function). After 10 min of simulated time, low αs are slightly overestimated

[6] The difference in scale between figures 2(A) and 5(C) is due to the normalization condition equation (64). For figure 2(A), $m/q = 110/1 = 110$; for figure 5(C), $m/q = 29/10 = 2.9$. Since $\|\boldsymbol{h}_1\| = \sqrt{m/q}$, the values of $\boldsymbol{h}_1$ for figure 5(C) will be $\sqrt{2.9/110} \approx 0.16$ times the values for figure 2(A). The result is an artefact of our insistence that the argument to the nonlinearity have variance one.

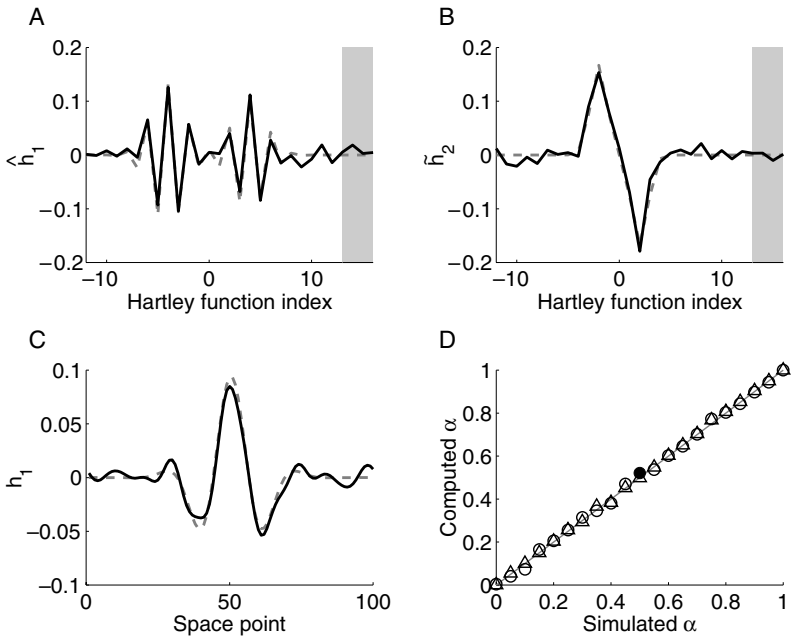

Figure 5. Results from simulations with a stimulus ensemble composed of Hartley functions. (A), (B) Panels are identical to figures 2(A), (B) except that the kernels are plotted in the space defined by the Hartley functions. Shown are kernels calculated from 10 min of simulated time with $\alpha = 0.5$. (C) The linear kernel of panel (A) transformed into spatial coordinates. The responses to blank stimuli were ignored in the transformation. (D) Comparison of simulated and computed QIN α for all simulations with Hartley function stimuli. Results from 10 min (circles) and 1 h (triangles) of stimulated time are shown. The filled circle indicates the simulation of the previous panels.

and high αs are slightly underestimated, but this deviation disappears after 1 h of simulated time.

5. Discussion

We have demonstrated a method to separate the linear (odd-order) and quadratic (even-order) contributions to neuronal response as well as measure their relative contribution to the response. The method is robust to deviations from the assumptions required for its derivation. In particular, the assumptions of an exponential nonlinearity and Gaussian input to the nonlinearity can be relaxed with only minor degradation of the results.

5.1. Local squaring of the model

The linear–quadratic–nonlinear model (equation (2)) was designed to capture the manner in which complex cells in the primary visual cortex can respond independently of contrast sign. The local (pixel-by-pixel) squaring of stimulus values is a phenomenological model of this sign independence. We chose this simple representation because it can be completely reconstructed from realistic experiments. The results of the analysis have a model-independent interpretation (see section 3.5) that do not depend on a neural basis for squaring. Nonetheless, Archie and Mel [2] demonstrate how active dendrites could act like squaring subunits. Such squaring is more general than the squaring of individual stimulus components in equation (2).

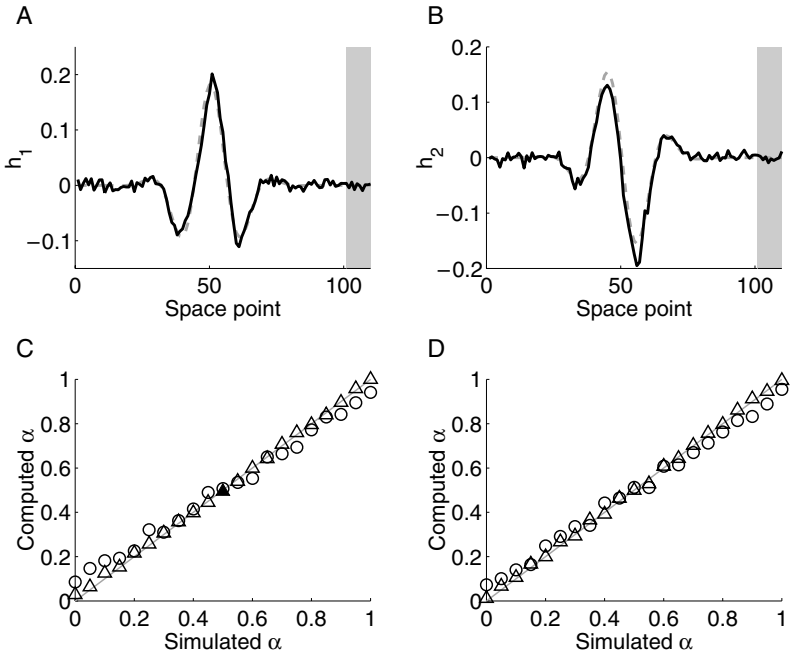

Figure 6. Results with power law and error function nonlinearities. (A), (B) Panels are identical to figures 2(A), (B). Shown are kernels calculated from 1 h of simulated time with $\alpha = 0.5$ and power law nonlinearity. (C), (D) Comparison of simulated and computed QIN α for power law (C) and error function (D) simulations. Results from 10 min (circles) and 1 h (triangles) of stimulated time are shown. The filled triangle in panel (C) indicates the simulation of the previous panels.

In fact, given that the stimulus (when viewed as random dots) includes only the values 1 and -1 (when nonzero), the distinction between linear and quadratic is equivalent to the distinction between odd-order and even-order functions of the input. The squaring enters the analysis only to remove the negative sign from -1. This fact makes the QIN truly a measure of the contrast sign independence. For this reason, one might view the QIN as a misnomer. We use the quadratic terminology and formulation of model (2) to allow application in principle to other stimuli.

5.2. Comparison to other correlation methods

In the case when $\alpha = 0$, equation (2) is the classical linear–nonlinear model. Correlation methods are commonly used to recover the linear kernel $\boldsymbol{h}_1$ in this model [1, 3, 5–7, 9, 12, 17, 20, 24]. When the stimulus is a sequence of random orthogonal vectors, Ringach *et al* [20] showed that $\boldsymbol{h}_1$ is proportional to the stimulus–spike correlation when the nonlinearity is a simple linear threshold. With the exponential nonlinearity assumed in our analysis, this relationship is no longer exactly correct. Since our results are applicable even for $\alpha = 0$, they demonstrate that $\boldsymbol{h}_1$ is instead proportional to $\boldsymbol{v}_1$ (equation (27)). In simulations, we observed only minor differences between the two estimates of $\boldsymbol{h}_1$ when $\alpha = 0$. The advantage of the proposed method is that it eliminates any sign-independent effects from $\boldsymbol{h}_1$ in case they exist.

When $\alpha = 1$, the neuron responds to the stimulus in a completely sign-independent fashion. Consequently, $E\{X^i_j R\} = 0$ and equation (28) for $\boldsymbol{v}_2$ becomes

$$v^i_{2,j} = \log\left(\frac{E\{(X^i_j)^2 R\}}{E\{K^i\}}\right).$$

When the stimulus ensemble is composed of two-dimensional Hartley functions, then this formula is nearly identical to that used by Ringach *et al* [18] to study the response of visual cortex neurons to random gratings. Ringach *et al* simply average over additional phases of the gratings. Ringach *et al* also provide a justification for the logarithm based on the detection of deviations from the baseline, which can be viewed as a justification for the exponential nonlinearity.

5.3. Comparison to other measures of linearity

The QIN α is an estimate of the linearity of visual cortex neurons, ranging from the idealized simple cell to the idealized complex cell. In this sense, it is similar to the F_1/F_0 ratio, a measure of the linearity of spatio-temporal summation, which was proposed by a number of investigators as an appropriate measure of cell complexity [4, 11, 14, 22]. F_1/F_0 represents the ratio between the amplitude of the first harmonic response and the mean spike rate when cortical neurons are stimulated with drifting sinusoidal gratings.

The relationship between the QIN and the F_1/F_0 ratio are studied extensively in [15], both through simulations and through analysis of the response of neurons in the primary visual cortex. Those results indicate that the QIN is a better indicator of nonlinearity in receptive field organization because the F_1/F_0 ratio is sensitive to the nonlinearity of the spike threshold [13]. The effect of the threshold nonlinearity confounds the effect of receptive field nonlinearity in the F_1/F_0 ratio, leading to difficulties in interpreting F_1/F_0. The QIN, on the other hand, is insensitive to changes in the nonlinearity $g(\cdot)$ (cf figure 6) and thus the spiking threshold.

The simulations of [15] indicate that the QIN is a robust measure of overlap between subfields of a traditional complex cell model, similar to the degree of overlap between ON and OFF subfields proposed by Hubel and Wiesel [8] as a method to classify simple and complex cells. In this way, the QIN is similar to the overlap index proposed by Schiller *et al* [21] and recently used by Kagan *et al* [10]. The advantages of the QIN include not only its demonstrated robustness to the spike threshold, but also its applicability to a broad class of spatio-temporally rich stimuli, its sensitivity to subthreshold effects and its foundation on a explicit mathematical model. For these reasons, the QIN is a robust tool that one can confidently use to analyse receptive field properties of neurons in the primary visual cortex and other primary sensory areas.

Acknowledgments

The author thanks Dario Ringach and an anonymous reviewer for insightful suggestions and comments. This work was supported by an NSF Mathematical Sciences Postdoctoral Research Fellowship.

Appendix A. Mathematical details

A.1. Mean and variance calculations

In this section, we calculate the mean and variance of the random variables that will be approximated as Gaussian. First we calculate the statistics of the components of Y_R:

$$\begin{aligned} Y_R &= \sqrt{1-\alpha}\boldsymbol{h}_1 \cdot \boldsymbol{X} + \sqrt{\alpha}\boldsymbol{h}_2 \cdot \boldsymbol{X}^2 \\ &= \sqrt{1-\alpha}Y_{R1} + \sqrt{\alpha}Y_{R2}. \end{aligned}$$

Recall that a particular realization of the input is determined by the choice of dot positions $\boldsymbol{J}$ and dot signs $\boldsymbol{B}$. It can be written as

$$\boldsymbol{X} = \sum_{i=1}^{n} B^i \boldsymbol{e}^i_{J^i}, \tag{34}$$

where $\boldsymbol{e}^i_j$ indicates a dot at position j and time i (see equation (6)). With this notation, the component Y_{R1} is

$$Y_{R1} = \boldsymbol{h}_1 \cdot \boldsymbol{X} = \sum_{k=1}^{n} B^k h^k_{1,J^k}, \tag{35}$$

and its square is

$$(Y_{R1})^2 = \sum_{k=1}^{n}\sum_{l=1}^{n} B^k B^l h^k_{1,J^k} h^l_{1,J^l}. \tag{36}$$

The average value of Y_{R1} is the sum over all $(2m)^2$ possible realizations of the recent stimulus $\boldsymbol{X}$ divided by $(2m)^2$:

$$E\{Y_{R1}\} = \frac{1}{(2m)^n} \sum_{\boldsymbol{J}\in\mathcal{J}^n_m} \sum_{\boldsymbol{B}\in\mathcal{B}^n} \left(\sum_{k=1}^{n} B^k h^k_{1,J^k} \right), \tag{37}$$

where $\mathcal{J}^n_m$ and $\mathcal{B}^n$ indicate the sets of all possible sequences of dot positions and signs. Since each term $B^k h^k_{1,J^k}$ can be paired with its opposite $-B^k h^k_{1,J^k}$, all terms in the expression for $E\{Y_{R1}\}$ cancel out. The mean value of Y_{R1} is

$$\mu_{R1} = E\{Y_{R1}\} = 0. \tag{38}$$

In calculating $E\{(Y_{R1})^2\}$, terms with $B^k B^l$, $k \neq l$ can be paired with the corresponding term with $-B^k B^l$ and be cancelled out. Only the terms with $k = l$ survive since $(B^k)^2 = 1$ for either sign of B^k:

$$\begin{aligned} E\{(Y_{R1})^2\} &= \frac{1}{(2m)^n} \sum_{\boldsymbol{J}\in\mathcal{J}^n_m} \sum_{\boldsymbol{B}\in\mathcal{B}^n} \sum_{k=1}^{n}\sum_{l=1}^{n} B^k B^l h^k_{1,J^k} h^l_{1,J^l} \\ &= \frac{1}{(2m)^n} \sum_{\boldsymbol{J}\in\mathcal{J}^n_m} \sum_{\boldsymbol{B}\in\mathcal{B}^n} \sum_{k=1}^{n} (h^k_{1,J^k})^2. \\ &= \frac{1}{m^n} \sum_{\boldsymbol{J}\in\mathcal{J}^n_m} \sum_{k=1}^{n} (h^k_{1,J^k})^2. \end{aligned} \tag{39}$$

For any given j and k with $1 \leqslant j \leqslant m$ and $1 \leqslant k \leqslant n$, exactly m^{n-1} vectors in $\mathcal{J}^n_m$ have $J^k = j$ (since only one component of the vector $\boldsymbol{J}$ is specified, the remaining $n-1$ components can be chosen freely). Each term $(h^k_{1,j})^2$ appears exactly m^{n-1} times in the final sum of equation (39) and $E\{(Y_{R1})^2\}$ simplifies to

$$\begin{aligned} E\{(Y_{R1})^2\} &= \frac{1}{m} \sum_{j=1}^{m}\sum_{k=1}^{n} (h^k_{1,j})^2 \\ &= \frac{\|\boldsymbol{h}_1\|^2}{m}, \end{aligned} \tag{40}$$

where $\|\cdot\|$ is the standard vector norm. The variance of Y_{R1} is then

$$\sigma^2_{R1} = E\{(Y_{R1})^2\} - (E\{Y_{R1}\})^2 = \frac{\|\boldsymbol{h}_1\|^2}{m},$$

which is equation (19).

We repeat the same procedure for Y_{R2}. In this case, since everything is positive due to squaring, we do not have the cancellations we did with Y_{R1}. For a particular value of the input as in equation (34), the local squaring in $\boldsymbol{X}^2$ simply removes all B^i (since $(B^i)^2 = 1$):

$$\boldsymbol{X}^2 = \sum_{i=1}^{n} \boldsymbol{e}^i_{J^i}. \tag{41}$$

This means that

$$Y_{R2} = \boldsymbol{h}_2 \cdot \boldsymbol{X}^2 = \sum_{k=1}^{n} h^k_{2,J^k} \tag{42}$$

and

$$(Y_{R2})^2 = \sum_{k=1}^{n} \sum_{l=1}^{n} h^k_{2,J^k} h^l_{2,J^l}. \tag{43}$$

The average value of Y_{R2} is the sum over all $(2m)^2$ possible values of $\boldsymbol{X}$ divided by $(2m)^2$. In this case, since the terms do not depend on the B^k, we can immediately sum over all 2^n possible $\boldsymbol{B} \in \mathcal{B}^n$:

$$\begin{aligned} E\{Y_{R2}\} &= \frac{1}{(2m)^n} \sum_{\boldsymbol{J} \in \mathcal{J}^n_m} \sum_{\boldsymbol{B} \in \mathcal{B}^n} \left(\sum_{k=1}^{n} h^k_{2,J^k} \right) \\ &= \frac{1}{m^n} \sum_{\boldsymbol{J} \in \mathcal{J}^n_m} \sum_{k=1}^{n} h^k_{2,J^k}. \end{aligned} \tag{44}$$

By a similar argument as above, each $h^k_{2,j}$ appears in the sum m^{n-1} times. $E\{Y_{R2}\}$, which we denote by μ_{R2}, simplifies to

$$\begin{aligned} \mu_{R2} = E\{Y_{R2}\} &= \frac{1}{m} \sum_{j=1}^{m} \sum_{k=1}^{n} h^k_{2,j} \\ &= \sum_{k=1}^{n} \bar{h}^k_2, \end{aligned} \tag{45}$$

(equation (11)) where $\bar{h}^k_2$ is the average value of $\boldsymbol{h}_2$ at time point k:

$$\bar{h}^k_2 = \frac{1}{m} \sum_{j=1}^{m} h^k_{2,j}. \tag{46}$$

The average of $(Y_{R2})^2$ (equation (43)) also does not depend on the B^k:

$$\begin{aligned} E\{(Y_{R2})^2\} &= \frac{1}{m^n} \sum_{\boldsymbol{J} \in \mathcal{J}^n_m} \sum_{k=1}^{n} \sum_{l=1}^{n} h^k_{2,J^k} h^l_{2,J^l} \\ &= \frac{1}{m^n} \sum_{\boldsymbol{J} \in \mathcal{J}^n_m} \sum_{k=1}^{n} (h^k_{2,J^k})^2 + \frac{1}{m^n} \sum_{\boldsymbol{J} \in \mathcal{J}^n_m} \sum_{k=1}^{n} \sum_{\substack{l=1 \\ l \neq k}}^{n} h^k_{2,J^k} h^l_{2,J^l} \\ &= \frac{1}{m} \sum_{k=1}^{n} \sum_{j=1}^{m} (h^k_{2,j})^2 + \frac{1}{m^2} \sum_{\substack{k,l=1 \\ l \neq k}}^{n} \sum_{i,j=1}^{m} h^k_{2,j} h^l_{2,i}. \end{aligned} \tag{47}$$

We separated out terms where $k = l$. These terms appear m^{n-1} times in the large sum. For $k \neq l$, requiring $J^k = j$ and $J^l = i$ leaves $n - 2$ components of $\boldsymbol{J}$ to be chosen freely. Thus, the terms $h^k_{2,j} h^l_{2,i}$ appear m^{n-2} times in the large sum.

We rewrite $E\{(Y_{R2})^2\}$ in terms of $\bar{h}_2^k$:

$$E\{(Y_{R2})^2\} = \frac{1}{m}\sum_{k=1}^{n}\sum_{j=1}^{m}(h_{2,j}^k)^2 + \sum_{\substack{k,l=1\\ l\neq k}}^{n}\left(\frac{1}{m}\sum_{j=1}^{m}h_{2,j}^k\right)\left(\frac{1}{m}\sum_{i=1}^{m}h_{2,i}^l\right)$$
$$= \frac{1}{m}\sum_{k=1}^{n}\sum_{j=1}^{m}(h_{2,j}^k)^2 + \sum_{\substack{k,l=1\\ l\neq k}}^{n}\bar{h}_2^k\bar{h}_2^l. \tag{48}$$

Combining this with equation (45), the variance of Y_{R2} is

$$\sigma_{R2}^2 = E\{(Y_{R2})^2\} - \mu_{R2}^2$$
$$= \frac{1}{m}\sum_{k=1}^{n}\sum_{j=1}^{m}(h_{2,j}^k)^2 + \sum_{\substack{k,l=1\\ l\neq k}}^{n}\bar{h}_2^k\bar{h}_2^l - \sum_{k,l=1}^{n}\bar{h}_2^k\bar{h}_2^l$$
$$= \sum_{k=1}^{n}\left[\frac{1}{m}\sum_{j=1}^{m}(h_{2,j}^k)^2 - (\bar{h}_2^k)^2\right]. \tag{49}$$

The term for each time point k is the variance of the values of $\boldsymbol{h}_2$ for that time. We define a function $\sigma_t(\boldsymbol{v})$ acting on vectors $\boldsymbol{v}$ the same size as $\boldsymbol{h}_2$ by

$$\sigma_t(\boldsymbol{v}) = \left\{\sum_{k=1}^{n}\left[\frac{1}{m}\sum_{j=1}^{m}(v_j^k)^2 - \left(\frac{1}{m}\sum_{j=1}^{m}v_j^k\right)^2\right]\right\}^{1/2}. \tag{50}$$

Then $\sigma_t^2(\boldsymbol{v})$ is the sum of the variance of each time slice of $\boldsymbol{v}$. In terms of this function, the variance of Y_{R2} is simply

$$\sigma_{R2}^2 = \sigma_t^2(\boldsymbol{h}_2)$$

which is equation (20).

Since Y_{R1} and Y_{R2} are clearly not independent, one would not expect to be able to add the variances to obtain the variance of $Y_R = \sqrt{1-\alpha}Y_{R1} + \sqrt{\alpha}Y_{R2}$. However, the covariance between Y_{R1} and Y_{R2} is zero:

$$E\{Y_{R1}Y_{R2}\} - E\{Y_{R1}\}E\{Y_{R2}\} = 0 - 0 = 0.$$

$E\{Y_{R1}Y_{R2}\} = 0$ because each term in $Y_{R1}Y_{R2}$ is linear in the binary variables B^k. Each term is cancelled by an equivalent term of opposite sign, just as in the calculation of $E\{Y_{R1}\}$. Consequently, the variance of Y_R is simply given by the sum of the two component variances

$$\sigma_R^2 = (1-\alpha)\sigma_{R1}^2 + \alpha\sigma_{R2}^2 = 1.$$

For the calculation of the stimulus–spike correlation, we define the two components of Y_{X^iR} just like we did for Y_R:

$$Y_{X^iR} = \sqrt{1-\alpha}Y_{X^iR1} + \sqrt{\alpha}Y_{X^iR2} \tag{51}$$

where

$$Y_{X^iR1} = \sum_{\substack{k=1\\ k\neq i}}^{n}\sum_{j=1}^{m}h_{1,j}^i X_j^i \tag{52}$$

and

$$Y_{X^iR2} = \sum_{\substack{k=1\\ k\neq i}}^{n}\sum_{j=1}^{m}h_{2,j}^i (X_j^i)^2. \tag{53}$$

The mean and variance for these quantities can be calculated exactly as for Y_R. The only difference is that time point i is omitted so that the averages are over $(2m)^{n-1}$ stimulus possibilities. The expressions for the mean $\mu_{X^i R}$ and variance $\sigma^2_{X^i R}$ of $Y_{X^i R}$ differ from the expressions for Y_R only by the omission of time point i:

$$\begin{aligned}
\mu_{X^i R} &= \sqrt{1-\alpha}\mu_{X^i R1} + \sqrt{\alpha}\mu_{X^i R2} \\
\mu_{X^i R1} &= 0 \\
\mu_{X^i R2} &= \sum_{\substack{k=1 \\ k\neq i}}^{n} \bar{h}_2^k \\
\sigma^2_{X^i R} &= (1-\alpha)\sigma^2_{X^i R1} + \alpha\sigma^2_{X^i R2} \\
\sigma^2_{X^i R1} &= \frac{1}{m}\sum_{j=1}^{m}\sum_{\substack{k=1 \\ k\neq i}}^{n}(h^k_{1,j})^2 \\
\sigma^2_{X^i R2} &= \sum_{\substack{k=1 \\ k\neq i}}^{n}\left[\frac{1}{m}\sum_{j=1}^{m}(h^k_{2,j})^2 - (\bar{h}^k_2)^2\right].
\end{aligned}$$

A.2. General orthonormal stimuli

To generalize the analysis to random sequences of general orthonormal stimuli in the set $\{\hat{\boldsymbol{e}}_j\}$, we simply perform the analysis in the coordinates determined by the stimulus elements. Given the orthonormal condition (equation (3)), we can let the vectors $\{\hat{\boldsymbol{e}}_j\}$ be the coordinate axes and simply identify $\hat{\boldsymbol{e}}_j$ with the m-dimensional vector that is zero except for the jth component.

This identification is an invertible transform analogous to the discrete Fourier transform. If $\boldsymbol{v}$ is a vector in the span of $\{\hat{\boldsymbol{e}}_j\}$, then its transform $\hat{\boldsymbol{v}}$ is defined by

$$\hat{v}_j = \langle \boldsymbol{v}, \hat{\boldsymbol{e}}_j \rangle = \sum_k v_k \hat{e}_{j,k} \tag{54}$$

where $\hat{e}_{j,k}$ are the components of $\hat{\boldsymbol{e}}_j$. The inverse transform is

$$v_k = \sum_{j=1}^{m} \hat{v}_j \hat{e}_{j,k}. \tag{55}$$

Through this transform, the sequence of images is mathematically identical to the sequence of dots used in the above analysis. Each presentation of a black or white dot on a grey background now must be interpreted as the coefficients of each of the orthonormal vectors.

No further modifications would be necessary if one were comfortable assuming that the model given by equation (2) was acceptable when each stimulus component X^i_j is the coefficient of the jth orthonormal vector at time i. However, a model that was defined in physical space makes more sense from a neuroscience perspective. Fortunately, for our stimulus, a model given by equation (2) in physical space corresponds to a model of the same form in the space of orthonormal vector coefficients.

That the linear term $\boldsymbol{h}_1 \cdot \boldsymbol{X}$ corresponds to an equivalent term $\hat{\boldsymbol{h}}_1 \cdot \hat{\boldsymbol{X}}$ in coefficient space is a well known mathematical fact. For this discussion, we look just at a single time point i and let the i superscript on a vector denote this time slice. Then, $\hat{\boldsymbol{h}}^i_1$ is simply the transform of $\boldsymbol{h}^i_1$ just as the $\hat{\boldsymbol{X}}^i$ is the transform of $\boldsymbol{X}^i$:

$$\boldsymbol{h}^i_1 \cdot \boldsymbol{X}^i = \sum_k h^i_{1,k} X^i_k = \sum_k h^i_{1,k} \sum_j \hat{X}^i_j \hat{e}_{j,k}$$

$$= \sum_j \hat{X}^i_j \sum_k h^i_{1,k} \hat{e}_{j,k} = \sum_j \hat{h}^i_{1,j} \hat{X}^i_j$$
$$= \hat{\boldsymbol{h}}^i_1 \cdot \hat{\boldsymbol{X}}^i. \tag{56}$$

Repeating this procedure for each time slice, we determine that $\boldsymbol{h}_1 \cdot \boldsymbol{X} = \hat{\boldsymbol{h}}_1 \cdot \hat{\boldsymbol{X}}$, where the notation $\hat{\boldsymbol{v}}$ indicates taking the transform (54) for each time slice. One can compute $\hat{\boldsymbol{h}}_1$ by treating the stimulus as $\hat{\boldsymbol{X}}$, reducing the problem to the case of dots. Then the kernel $\boldsymbol{h}_1$ is simply the inverse transform of $\hat{\boldsymbol{h}}_1$. Since $\|\boldsymbol{h}_1\| = \|\hat{\boldsymbol{h}}_1\|$, applying the normalization of equation (22) to $\hat{\boldsymbol{h}}_1$ is equivalent to applying it to $\boldsymbol{h}_1$; computing $\boldsymbol{h}_1$ in the transform domain does not affect the calculation of the other parameters.

The motivation for presenting these classical results is to compare them with the quadratic term. Due to the nonlinear squaring, one does not simply obtain the transformed version of $\boldsymbol{h}_2$. Fortunately, due to the structure of the stimulus, we do obtain a similar term $\tilde{\boldsymbol{h}}_2 \cdot \hat{\boldsymbol{X}}^2$ that corresponds to $\boldsymbol{h}_2 \cdot \boldsymbol{X}^2$, where the notation $\tilde{\boldsymbol{h}}_2$ is defined below.

Since only one image is presented at each time slice, the coefficient of only one component $\hat{X}^i_j$ can be nonzero for a given i. The product $\hat{X}^i_j \hat{X}^i_p$ must be zero for $p \neq j$. This fact simplifies the dot product of the time slices when expressed in terms of the transforms:

$$\boldsymbol{h}^i_2 \cdot (\boldsymbol{X}^i)^2 = \sum_k h^i_{2,k} (X^i_k)^2 = \sum_k h^i_{2,k} \left(\sum_j \hat{X}^i_j \hat{e}_{j,k} \right)^2$$
$$= \sum_j (\hat{X}^i_j)^2 \sum_k h^i_{2,k} (\hat{e}_{j,k})^2 + \sum_j \sum_{p \neq j} \hat{X}^i_j \hat{X}^i_p \sum_k h^i_{2,k} \hat{e}_{j,k} \hat{e}_{p,k}$$
$$= \sum_j \tilde{h}^i_{2,j} (\hat{X}^i_j)^2 = \tilde{\boldsymbol{h}}^i_2 \cdot (\hat{\boldsymbol{X}}^i)^2. \tag{57}$$

We have defined another transform in terms of squared values of the orthonormal vectors as

$$\tilde{v}_j = \langle \boldsymbol{v}, \hat{\boldsymbol{e}}^2_j \rangle = \sum_k v_k (\hat{e}_{j,k})^2. \tag{58}$$

Repeating this procedure for each time slice, we determine that $\boldsymbol{h}_2 \cdot \boldsymbol{X}^2 = \tilde{\boldsymbol{h}}_2 \cdot \hat{\boldsymbol{X}}^2$, where the notation $\tilde{\boldsymbol{v}}$ indicates taking the transform (58) for each time slice.

Unfortunately, the vectors $\{\hat{\boldsymbol{e}}^2_j\}$ will, in general, be linearly dependent. The transform (58) will not be invertible; it is impossible to recover $\boldsymbol{h}_2$ from $\tilde{\boldsymbol{h}}_2$. One could compute the projection of each time slice $\boldsymbol{h}^i_2$ onto the subspace spanned by the $\{\hat{\boldsymbol{e}}^2_j\}$ (that projection is the only part of $\boldsymbol{h}^i_2$ that affects $\boldsymbol{h}_2 \cdot \boldsymbol{X}^2$), but its interpretation would be unclear without first understanding the subspace.

Although the nature of the stimulus prevents a reconstruction of $\boldsymbol{h}_2$, the fact that $\boldsymbol{h}^i_2 \cdot (\boldsymbol{X}^i)^2 = \tilde{\boldsymbol{h}}^i_2 \cdot (\hat{\boldsymbol{X}}^i)^2$ allows the analysis to be performed in the transformed space of coefficients even when starting with model (2) in physical space. If we use the normalization

$$\sigma_t^2(\tilde{\boldsymbol{h}}_2) = 1 \tag{59}$$

rather than equation (22), then $\boldsymbol{h}_2 \cdot \boldsymbol{X}$ will have variance one, and the calculation of the other parameters will be the same as in the case of random dots.

A.3. Slower stimuli

We generalize the analysis to the case when the stimulus $\boldsymbol{X}$ changes only every q time points. This slowing by a factor of q increases the effective power or variance of the stimulus by a factor of q if the stimulus values are held constant (this is the same phenomenon as when changing the temporal discretization of white noise approximations).

Since we normalize the kernels so that $\boldsymbol{h}_1 \cdot \boldsymbol{X}$ and $\boldsymbol{h}_2 \cdot \boldsymbol{X}^2$ have unit variance, we must scale the kernels by a factor of $1/\sqrt{q}$ (compare equations (64) and (67) with (22) and (29)). We also scale the parameter β by $1/\sqrt{q}$.

This scaling is the only change from the original case. In the rest of this appendix, we hammer out the details to demonstrate that these claims are indeed true.

The slow stimulus rate will effectively smooth measured stimulus–spike correlations and prevent resolving temporal structures finer than the stimulus rate. Even though we will compute $\boldsymbol{h}_1$ and $\boldsymbol{h}_2$ at every time point, we assume that they change insignificantly over one stimulus frame so that we can shift the temporal index of the order of q units of time as needed to simplify the calculations.

We first shift temporal indices so that the given time is at the end of a stimulus presentation. Then, we increase n to the nearest multiple of q and denote by $\tilde{n} = n/q$ the number of previous stimulus frames that affect a neuron's response. Our set of possible stimuli (4) is replaced by the $(2m)^{\tilde{n}}$ possibilities

$$\boldsymbol{X} \in \left\{ \sum_{i=0}^{\tilde{n}-1} \sum_{p=1}^{q} B^i \boldsymbol{e}_{J^i}^{qi+p} : \boldsymbol{J} \in \mathcal{J}_m^{\tilde{n}} \,\&\, \boldsymbol{B} \in \mathcal{B}^{\tilde{n}} \right\}. \tag{60}$$

The calculation of $E\{R\}$ proceeds as in section 3.3.1. The only departure from that derivation is the calculation of $E\{(Y_{R1})^2\}$ and $E\{(Y_{R2})^2\}$. The variance of Y_{R1} is

$$\begin{aligned}
\sigma_{R1}^2 = E\{(Y_{R1})^2\} &= \frac{1}{(2m)^{\tilde{n}}} \sum_{\boldsymbol{J} \in \mathcal{J}_m^{\tilde{n}}} \sum_{\boldsymbol{B} \in \mathcal{B}^{\tilde{n}}} \sum_{k,l=0}^{\tilde{n}-1} \sum_{p,r=1}^{q} B^k B^l h_{1,J^k}^{qk+p} h_{1,J^l}^{ql+r} \\
&= \frac{1}{m^{\tilde{n}}} \sum_{\boldsymbol{J} \in \mathcal{J}_m^{\tilde{n}}} \sum_{k=0}^{\tilde{n}-1} \sum_{p,r=1}^{q} h_{1,J^k}^{qk+p} h_{1,J^k}^{qk+r} \\
&= \frac{q}{m^{\tilde{n}}} \sum_{\boldsymbol{J} \in \mathcal{J}_m^{\tilde{n}}} \sum_{k=0}^{\tilde{n}-1} \sum_{p=1}^{q} \left(h_{1,J^k}^{qk+p}\right)^2 \\
&= \frac{q}{m} \|h_{1,j}^k\|^2
\end{aligned} \tag{61}$$

where the only new feature is approximating h_{1,J^k}^{qk+r} by h_{1,J^k}^{qk+p} by shifting the temporal index of the order of q units of time.

The expected value of Y_{R2}^2 is

$$\begin{aligned}
E\{(Y_{R2})^2\} &= \frac{1}{m^{\tilde{n}}} \sum_{\boldsymbol{J} \in \mathcal{J}_m^{\tilde{n}}} \sum_{k,l=0}^{\tilde{n}-1} \sum_{p,r=1}^{q} h_{2,J^k}^{qk+p} h_{2,J^l}^{ql+r} \\
&= \frac{1}{m^{\tilde{n}}} \sum_{\boldsymbol{J} \in \mathcal{J}_m^{\tilde{n}}} \sum_{k=0}^{\tilde{n}-1} \sum_{p,r=1}^{q} h_{2,J^k}^{qk+p} h_{2,J^k}^{qk+r} + \frac{1}{m^{\tilde{n}}} \sum_{\boldsymbol{J} \in \mathcal{J}_m^{\tilde{n}}} \sum_{\substack{k,l=0 \\ l \neq k}}^{\tilde{n}-1} \sum_{p,r=1}^{q} h_{2,J^k}^{qk+p} h_{2,J^l}^{ql+r} \\
&= \frac{q}{m} \sum_{k=1}^{n} \sum_{j=1}^{m} (h_{2,j}^k)^2 + \sum_{k=1}^{n} \left(\sum_{l=1}^{n} \bar{h}_2^k \bar{h}_2^l - q(\bar{h}_2^k)^2 \right).
\end{aligned} \tag{62}$$

We approximated h_{2,J^k}^{qk+r} by h_{2,J^k}^{qk+p} in the first sum before making the counting argument to arrive at the last step. For the second sum, we shifted the temporal indices of the missing terms to k after making the counting argument. The variance of Y_{R2} is then

$$\sigma_{R2}^2 = q \sum_{k=1}^{n} \left[\frac{1}{m} \sum_{j=1}^{m} (h_{2,j}^k)^2 - (\bar{h}_2^k)^2 \right] = q \sigma_t^2(\boldsymbol{h}_2). \tag{63}$$

Equations (61) and (63) show that the effective power of the stimulus is proportional to the frame interval q. The firing rate will increase as the stimulus is slowed even for a model with no preferred temporal frequency built in. We mask this dependence on the stimulus frequency by requiring the nonlinearity argument to have variance one. The kernels are simply divided by $\sqrt{q}$ as compared to the original derivations, the normalization conditions becoming

$$\|\boldsymbol{h}_1\| = \sqrt{m/q} \qquad \text{and} \qquad \sigma_t(\boldsymbol{h}_2) = 1/\sqrt{q}. \tag{64}$$

In the calculation of the stimulus–spike correlations, the $h^i_{1,j}$ and $h^i_{2,j}$ terms are multiplied by q. For example, in the correlation of the spikes with the positive stimuli, we know that if $X^i_j = 1$, then $X^k_j = 1$ for q temporal indices k around i. We pull out each of these terms from $Y_{X^i R}$ and approximate their temporal index as i, yielding $q\sqrt{1-\alpha}h^i_{1,j} + q\sqrt{\alpha}h^i_{2,j} + Y_{X^i R}$ as the nonlinearity argument. (The mean $\mu_{X^i R}$ and variance $\sigma^2_{X^i R}$ of $Y_{X^i R}$ change since more terms are removed, but they drop out of the final calculations.) We arrive at the modified expression for the stimulus–spike correlations

$$E\{X^i_j(X^i_j \pm 1)R\} = \frac{1}{m\sigma_{X^i R}\sqrt{2\pi}} \times \int g(y)\exp\left(-\frac{(y \mp q\sqrt{1-\alpha}h^i_{1,j} - q\sqrt{\alpha}h^i_{2,j} - \mu_{X^i R})^2}{2\sigma^2_{X^i R}}\right)\mathrm{d}y. \tag{65}$$

The analysis for an exponential nonlinearity in section 3.4.3 is identical except that equation (26) becomes

$$q\beta\left[\pm\sqrt{1-\alpha}h^i_{1,j} + \sqrt{\alpha}h^i_{2,j}\right] = \log\left(\frac{E\{X^i_j(X^i_j \pm 1)R\}}{E\{K^i\}}\right). \tag{66}$$

Given the new normalization condition (equation (64)), the expressions for the linear and quadratic kernels become

$$\boldsymbol{h}_1 = \frac{\sqrt{m}\boldsymbol{v}_1}{\sqrt{q}\|\boldsymbol{v}_1\|} \qquad \text{and} \qquad \boldsymbol{h}_2 = \frac{\boldsymbol{v}_2}{\sqrt{q}\sigma_t(\boldsymbol{v}_2)}, \tag{67}$$

so that

$$\|\boldsymbol{v}_1\|^2 = mq\beta^2(1-\alpha) \qquad \text{and} \qquad \sigma^2_t(\boldsymbol{v}_2) = q\beta^2\alpha.$$

The q falls out of the expression for the QIN so that equation (30) is still valid. The expression for β is multiplied by $1/\sqrt{q}$ and the expression for γ is unchanged.

Appendix B. Correcting finite sample effects

B.1. Bias reduction

Naive estimates of quantities, like $\|\boldsymbol{v}_1\|^2$ and $\sigma_t(\boldsymbol{v}_2)^2$, will be biased away from their true values due to noise. As discussed in [16], this bias increases with the size of the vectors. To obtain accurate results, steps must be taken to reduce the bias.

We derive a procedure for reducing the bias by viewing the estimates of the stimulus–spike statistics as random variables. The expected value of these random variables are the expected values assumed in our calculations. However, since the model parameters are nonlinear functions of those expected values, the variances and covariances of the random variables create bias in the nonlinear function of the random variables. In other words, the expected value of a nonlinear function of random variables is not equal to the nonlinear function of the expected value of the random variables.

Consider the estimation of $F(E\{A\}, E\{B\})$, where A and B are two random variables and $F(x, y)$ is some scalar nonlinear function. From a finite data set, one can estimate the random variables A and B (e.g. as sample means over the data set). However, even though A and B are unbiased estimates of $E\{A\}$ and $E\{B\}$, the estimate $F(A, B)$ will, in general, be a biased estimate of $F(E\{A\}, E\{B\})$:

$$E\{F(A, B)\} \neq F(E\{A\}, E\{B\}).$$

A major component of the bias is due to the covariance of the estimates A and B, $\operatorname{cov}(A, B)$. Computing the Taylor series, we obtain

$$\begin{aligned} F(A, B) \approx\ & F(E\{A\}, E\{B\}) + (A - E\{A\})\frac{\partial F}{\partial x}(E\{A\}, E\{B\}) \\ & + (B - E\{B\})\frac{\partial F}{\partial y}(E\{A\}, E\{B\}) + (A - E\{A\})^2\frac{\partial^2 F}{\partial x^2}(E\{A\}, E\{B\}) \\ & + (A - E\{A\})(B - E\{B\})\frac{\partial^2 F}{\partial x\,\partial y}(E\{A\}, E\{B\}) \\ & + (B - E\{B\})^2\frac{\partial^2 F}{\partial y^2}(E\{A\}, E\{B\}). \end{aligned}$$

Taking the expected values,

$$\begin{aligned} E\{F(A, B)\} \approx\ & F(E\{A\}, E\{B\}) + \sigma_A^2\frac{\partial^2 F}{\partial x^2}(E\{A\}, E\{B\}) \\ & + \operatorname{cov}(A, B)\frac{\partial^2 F}{\partial x\,\partial y}(E\{A\}, E\{B\}) + \sigma_B^2\frac{\partial^2 F}{\partial y^2}(E\{A\}, E\{B\}) \end{aligned}$$

where σ_A^2 is the variance of A and σ_B^2 is the variance of B.

To close the formula, we simply take the expected values out of the nonlinear function in the coefficients of the variance/covariance terms:

$$\begin{aligned} E\{F(A, B)\} \approx\ & F(E\{A\}, E\{B\}) + \sigma_A^2 E\left\{\frac{\partial^2 F}{\partial x^2}(A, B)\right\} \\ & + \operatorname{cov}(A, B)E\left\{\frac{\partial^2 F}{\partial x\,\partial y}(A, B)\right\} + \sigma_B^2 E\left\{\frac{\partial^2 F}{\partial y^2}(A, B)\right\}. \end{aligned}$$

Of course, this last step introduces bias of the same type we are attempting to eliminate. However, this error is approximately the same magnitude as the error we made in truncating the Taylor series. Assuming the Taylor series was expanded around small parameters, the quantity

$$F(A, B) - \sigma_A^2\frac{\partial^2 F}{\partial x^2}(A, B) - \operatorname{cov}(A, B)\frac{\partial^2 F}{\partial x\,\partial y}(A, B) - \sigma_B^2\frac{\partial^2 F}{\partial y^2}(A, B) \quad (68)$$

estimates $F(E\{A\}, E\{B\})$ with less bias than $F(A, B)$ alone.

For example, let A be an estimate of $E\{X_j^i(X_j^i + 1)R\}$ and B be an estimate of $E\{X_j^i(X_j^i - 1)R\}$ calculated from a data set. Then

$$v_{1,j}^k = \tfrac{1}{2}\log E\{A\} - \tfrac{1}{2}\log E\{B\}$$

and

$$(v_{1,j}^k)^2 = \tfrac{1}{4}(\log E\{A\})^2 - \tfrac{1}{2}\log E\{A\}\log E\{B\} + \tfrac{1}{4}(\log E\{B\})^2.$$

Each term is of the form $F(E\{A\}, E\{B\})$ and so should be estimated using expressions of the form (68). (To simplify the calculations, we ignored the second derivatives of the logarithm

because its contribution turned out to be inconsequential.) After simplification, the resulting estimator of $(v^k_{1,j})^2$ is

$$\left[\frac{1}{2}\log\left(\frac{C+D}{C-D}\right)\right]^2 - \frac{D\sigma_C^2 - 2CD\,\mathrm{cov}(C,D) + C\sigma_D^2}{(C^2-D^2)^2},$$

where $C = (A+B)/2$ and $D = (A-B)/2$ are estimates of $E\{(X^i_j)^2 R\}$ and $E\{X^i_j R\}$, respectively. In practice, we estimate the variances σ_C^2, σ_D^2 and covariance $\mathrm{cov}(C, D)$ by dividing the data set into trials and calculating estimates for each trial.

One can calculate estimators of each component of $\sigma_t^2(\boldsymbol{v}_2)$ using the same principle.

B.2. Correcting zero correlation estimates

When the nonlinearity $g(\cdot)$ is an exponential function, the expected value $E\{X^i_j(X^i_j \pm 1)R\}$ can never be exactly zero (although it can be arbitrarily close to zero). Even if model (2) with an exponential $g(\cdot)$ exactly represented the neural response, one may obtain zero estimates of $E\{X^i_j(X^i_j \pm 1)R\}$ from a finite data set. The estimate will be zero if no spikes happen to occur in the data set i units of time after presentation of a white or black dot at position j. In this case, our procedure to calculate the model parameters would require taking the logarithm of zero in equations (27) and (28).

One possible solution is to average over neighbouring time points. In some cases, however, we discovered we had to average over an unacceptably large number of points to obtain a nonzero value. To avoid such broad averaging, we developed the following procedure.

Let A be an estimate of $E\{X^i_j(X^i_j + 1)R\}$ and B be an estimate of $E\{X^i_j(X^i_j - 1)R\}$ calculated from a data set. Given our exponential form for $g(\cdot)$, we are assuming that $E\{A\} > 0$ and $E\{B\} > 0$. If we sample $A = 0$ or $B = 0$, then this sample must be below the true expected value. We arbitrarily set a constant ε and assume the sample A or B must be below the true mean by ε times their sample standard deviation σ_A or σ_B. Hence, we approximate the expected values by $E\{A\} \approx \varepsilon\sigma_A$ or $E\{B\} \approx \varepsilon\sigma_B$. We use the value $\varepsilon = 1$ (but see below).

If $A = 0$ or $B = 0$, then a straightforward estimate of its standard deviation would also be zero. (A and B are averages of nonnegative values. If the average is zero, all the terms composing the average are zero.) We compensate as follows. If both A and B are zero, then we approximate σ_A and σ_B as the average standard deviation observed from blank locations. If only A is zero, then we approximate σ_A as $\sigma_B/4$ (and approximate $\sigma_B \approx \sigma_A/4$ when $B = 0$).

Clearly, these choices are arbitrary. This approach can be viewed as a method of setting the exponential function to zero when its argument is small. To verify that this approach is not introducing significant error, we repeat the calculation of all model parameters when ε is doubled and when ε is halved. Large deviations in the estimates of the parameters as a result of these manipulations indicate that those estimates are not reliable. In all simulations, such deviations were minute and are therefore not represented in the figures. When applying this approach to experimental data, however, one must carefully examine deviations due to changing ε in order to throw out invalid results. One could also vary estimates of σ_A and σ_B independently and evaluate the resulting changes in parameters.

References

[1] Anzai A, Ohzawa I and Freeman R D 1999 Neural mechanisms for processing binocular information I. Simple cells *J. Neurophysiol.* **82** 891–908

[2] Archie K A and Mel B W 2000 A model for intradendritic computation of binocular disparity *Nat. Neurosci.* **3** 54–63

[3] Chichilnisky E J 2001 A simple white noise analysis of neural light responses *Network: Comput. Neural Syst.* **12** 199–213
[4] De Valois R L, Albrecht D G and Thorell L G 1982 Spatial frequency selectivity of cells in macaque visual cortex *Vis. Res.* **22** 545–9
[5] DeAngelis G C, Ohzawa I and Freeman R D 1993 Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. I. General characteristics and postnatal development *J. Neurophysiol.* **69** 1091–117
[6] DeAngelis G C, Ohzawa I and Freeman R D 1993 Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. II. Linearity of temporal and spatial summation *J. Neurophysiol.* **69** 1118–35
[7] Emerson R C, Korenberg M J and Citron M C 1989 Identification of intensive nonlinearities in cascade models of visual cortex and its relation to cell classification *Advanced Methods of Physiological System Modeling* ed V Z Marmarelis (New York: Plenum) pp 97–111
[8] Hubel D H and Wiesel T N 1962 Receptive fields, binocular interaction and functional architecture in the cat's visual cortex *J. Physiol.* **160** 106–54
[9] Jones J P and Palmer L A 1987 An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex *J. Neurophysiol.* **58** 1233–58
[10] Kagan I, Gur M and Snodderly D M 2002 Spatial organization of receptive fields of V1 neurons of alert monkeys: comparison with responses to gratings *J. Neurophysiol.* **88** 2557–74
[11] Maffei L and Fiorentini A 1973 The visual cortex as a spatial frequency analyser *Vis. Res.* **13** 1255–67
[12] Marmarelis P N and Marmarelis V Z 1978 *Analysis of Physiological Systems: The White Noise Approach* (New York: Plenum)
[13] Mechler F and Ringach D L 2002 On the classification of simple and complex cells *Vis. Res.* **42** 1017–33
[14] Movshon J A, Thompson I D and Tolhurst D J 1978 Spatial summation in the receptive fields of simple cells in the cat's striate cortex *J. Physiol. (Lond)* **283** 53–77
[15] Nykamp D Q and Ringach D L 2003 Quantifying the nonlinearity of receptive fields in primary visual cortex, submitted for publication
[16] Nykamp D Q and Ringach D L 2002 Full identification of a linear–nonlinear system via cross-correlation analysis *J. Vis.* **2** 1–11
[17] Reid R C, Victor J D and Shapley R M 1997 The use of m-sequences in the analysis of visual neurons: linear receptive field properties *Vis. Neurosci.* **14** 1015–27
[18] Ringach D L, Bredfeldt C E, Shapley R M and Hawken M J 2002 Suppression of neural responses to nonoptimal stimuli correlates with tuning selectivity in macaque V1 *J. Neurophysiol.* **87** 1018–27
[19] Ringach D L, Hawken M J and Shapley R 2002 Receptive field structure of neurons in monkey primary visual cortex revealed by stimulation with natural image sequences *J. Vis.* **2** 12–24
[20] Ringach D L, Sapiro G and Shapley R 1997 A subspace reverse-correlation technique for the study of visual neurons *Vis. Res.* **37** 2455–64
[21] Schiller P H, Finlay B L and Volman S F 1976 Quantitative studies of single-cell properties in monkey striate cortex. I. Spatiotemporal organization of receptive fields *J. Neurophysiol.* **39** 1288–319
[22] Skottun B C, De Valois R L, Grosof D H, Movshon J A, Albercht D G and Bonds A B 1991 Classifying simple and complex cells on the basis of response modulation *Vis. Res.* **31** 1079–86
[23] Theunissen F E, David S V, Singh N C, Hsu A, Vinje W E and Gallant J L 2001 Estimating spatio-temporal receptive fields of auditory and visual neurons from their responses to natural stimuli *Network: Comput. Neural Syst.* **12** 289–316
[24] Victor J D 1991 Asymptotic approach of generalized orthogonal functional expansions to Wiener kernels *Ann. Biomed. Eng.* **19** 383–99