Full identification of a linear-nonlinear system via cross-correlation analysis

Duane Q. Nykamp
Department of Mathematics, UCLA, Los Angeles, CA, USA

Dario L. Ringach
Departments of Psychology and Neurobiology and Brain Research Institute, UCLA, Los Angeles, CA, USA

A statistical model used extensively in vision research consists of a cascade of a linear operator followed by a static (memoryless) nonlinearity. Common applications include the measurement of simple-cell receptive fields in primary visual cortex and the modeling of human performance in various psychophysical tasks. It is well known that the front-end linear filter of the model can readily be recovered, up to a multiplicative constant, using reverse-correlation techniques. However, a full identification of the model also requires an estimation of the output nonlinearity. Here, we show that for a large class of static nonlinearities, one can obtain analytical expressions for the estimates. The technique works with both Gaussian and binary noise stimuli. The applicability of the method in physiology and psychophysics is demonstrated. Finally, the proposed technique is shown to converge much faster than the currently used linear-reconstruction method.

Introduction

A linear-nonlinear model consists of a cascade of a linear operator acting on the input in series with a static nonlinearity. In practice, the input (or stimulus) is a finite-dimensional vector, \( \mathbf{x} \), and the output (or response), \( r \), can be either 0 or 1. In this case, the model can be written simply as

\[ g(\mathbf{w}^T \mathbf{x}), \]

where \( g(\cdot) \) is a nonlinear function and the (column) vector \( \mathbf{w} \) represents the linear filtering operation. This equation can be considered a one-term projection pursuit regression model (Friedman & Stuetzle, 1981). Therefore, the optimization techniques developed for projection pursuit regression can be used to fit \( \mathbf{w} \) and \( g(\cdot) \) to the data. In the laboratory, however, the researcher has full control of the stimulus; therefore, as shown below, more efficient methods for estimating the parameters of the model can be developed.

Cross-correlation methods are commonly used to estimate the front-end linear filtering (the vector \( \mathbf{w} \)) in this model (Marmarelis & Marmarelis, 1978; Jones & Palmer, 1987; Emerson, Korenberg, & Citron, 1989; DeAngelis, Ohzawa, & Freeman, 1993a; DeAngelis, Ohzawa, & Freeman, 1993b; Reid, Victor, & Shapley, 1997; Ringach, Sapiro, & Shapley, 1997; Anzai, Ohzawa, & Freeman, 1999). A simple geometrical proof of the technique can be found in Chichilnisky (2001). In addition to estimating \( \mathbf{w} \), a full characterization of the model also requires that we estimate the shape of the nonlinearity.

The importance of estimating \( g(\cdot) \) has been somewhat neglected in the past [however, see Anzai et al., (1999) and Chichilnisky (2001)]. This is surprising because it is well known that the tuning properties of a neuron (such as its orientation or spatial frequency bandwidth) do not depend solely on its linear kernel; they can also be influenced by a static nonlinearity in the spike generation mechanism, represented by \( g(\cdot) \). Specifically, it has been suggested that both direction and orientation tuning can be sharpened significantly by thresholding or by an accelerating nonlinearity (Reid, Soodak, & Shapley, 1987; DeAngelis et al., 1993a,b; Anzai et al., 1999; Carandini & Ferster, 2000). Thus, estimating \( g(\cdot) \) is clearly important when comparing the tuning properties of the model to that of the data.

Similarly, cross-correlation methods are being used in visual psychophysics applications to recover the “classification images” used by subjects in two-alternative forced-choice (2AFC) tasks (Ahumada & Beard, 1998; Ringach, 1998; Gold, Murray, Bennett, & Sekuler, 2000). These methods show that performance is correlated, to some extent, with some physical attribute of the image. However, in order to quantitatively evaluate the fraction of human performance explained by the model, one needs to identify the static nonlinearity as well.

One strategy to identify the nonlinearity is to perform a two-step analysis (Anzai et al., 1999; Chichilnisky, 2001). First, measure the linear kernel using reverse correlation, then generate a scatter-plot of the linear prediction \( (\mathbf{w}^T \mathbf{x}) \) against the actual response \( r \) and smooth it to obtain an estimate of the static nonlinearity. Here we show that for a
large class of nonlinearities it is possible to write down a closed form solution for the estimates of the linear kernel and the static nonlinearity in one single step. The parameters of the nonlinearity are found by matching the input-output moments of the model to the data. We refer to this technique as the moment method. The calculation avoids computing the linear prediction altogether (which is a very time-consuming calculation) and can be updated recursively, as new data arrives. Furthermore, our results apply to both Gaussian and binary noise inputs.

**Model**

Let each component of the input be an independent normal (Gaussian) random variable with standard deviation $\sigma$. Alternatively, let each component be a binary random variable that takes the values $\sigma$ or $-\sigma$ with equal probability. As shown in Appendix A, if the linear kernel is normalized so that $\|w\| = 1$, the average output is

$$\bar{r} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} g(y) \exp\left( -\frac{y^2}{2\sigma^2} \right) dy.$$  \hspace{1cm} (2)

The correlation between the input and the output is proportional to the kernel $w$, $E[rx] = Cw$ \hspace{1cm} (3)

where $E[\cdot]$ denotes expected value and the constant of proportionality of $C$ is given by (Bussgang, 1952)

$$C = \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} g'(y) \exp\left( -\frac{y^2}{2\sigma^2} \right) dy.$$ \hspace{1cm} (4)

Thus, the linear filter $w$ can be recovered from $E[rx]$ because $w = E[rx]/E[rx]$, which is the usual reverse correlation result.

Through Equations 2 and 4 the measurement of the mean output, $\bar{r}$, and the proportionality constant, $C$, impose constraints on the nonlinearity $g(\cdot)$. The proposed moment method uses these measurements, which correspond to the first moment of the output $\bar{r}$ and the second input-output moment (cross-correlation), to specify the nonlinearity. For example, if one considers a class of nonlinearities specified by only two parameters, Equations 2 and 4 can be used to determine $g(\cdot)$. If, however, one considers a class of nonlinearities with more than two parameters, additional conditions must be provided to determine $g(\cdot)$. We demonstrate the procedure with some classes on nonlinearities used in both electrophysiology and psychophysics: half-rectification with threshold, a power law, an error function, and the Naka-Rushton equation.

**Half-Rectifier With Threshold Nonlinearity**

A linear-nonlinear system with a threshold half-rectifier nonlinearity has been one of the classical models of simple cells (Movshon, Thompson, & Tolhurst, 1978; Tolhurst & Heeger, 1997; Carandini & Ferster, 2000). The nonlinearity is described by, $g(y) = A[y - y_0]_+$, where $[x]_+$ is a half-rectifier such that $[x]_+ = x$ if $x > 0$ and is zero otherwise. In this case (see Appendix A), it can be shown that

$$\bar{r} = \frac{A\sigma}{\sqrt{2\pi}} \exp\left( -\frac{y_0^2}{2\sigma^2} \right) - \frac{y_0}{\sigma} C$$ \hspace{1cm} (5)

$$C = \frac{A\sigma^2}{2} \text{erfc}\left( \frac{y_0}{\sqrt{2\sigma^2}} \right)$$ \hspace{1cm} (6)

where erfc($x$) = $1 - \text{erf}(x)$ is the complementary error function and erf($x$) = $[2/\sqrt{\pi}]\int_{0}^{x} \exp(-t^2)dt$ is the error function. The system represented by Equations 5 and 6 has two measured variables, $\bar{r}$ and $C$, and two unknowns, $A$ and $y_0$. Thus, given the measurements, one can solve numerically the system to obtain estimates of $A$ and $y_0$.

**Power Law Nonlinearity**

This is an important example because power laws are used to model the output nonlinearity in simple cortical cells (Emerson et al, 1989; Albrecht & Geisler, 1991; Heeger, 1992; Anzai et al, 1999). Furthermore, they have been shown to fit the data slightly better than half-rectification (Tolhurst & Heeger, 1997). If we assume $g(y) = Ay^\beta$, then, as shown in Appendix A, we obtain

$$\bar{r} = \frac{1}{\sqrt{\pi}} \Gamma(\beta+1)/2$$ \hspace{1cm} (7)

$$C = \frac{A\beta}{\sqrt{\pi}} \sigma^{\beta+1/2} \Gamma(\beta/2)$$ \hspace{1cm} (8)

where the gamma function is given by $\Gamma(y) = \int_{0}^{\infty} v^{y-1}e^{-v}dv$. Once again, the system represented by Equations 7 and 8 has two measured variables, $\bar{r}$ and $C$, and two unknowns, $A$ and $\beta$. Thus, given the measurements one can solve numerically the system to obtain estimates of $A$ and $\beta$.

**Error Function Nonlinearity**

Another important case is when the nonlinearity can be represented by an error function:

$$g(y) = \frac{r_{max}}{2} \left[ 1 + \text{erf}\left( \frac{y - y_0}{\varepsilon\sqrt{2}} \right) \right].$$ \hspace{1cm} (9)

This class of nonlinearities is specified by three parameters: $r_{max}$, $y_0$, and $\varepsilon$. In Appendix A we derive the following specific forms of Equations 2 and 4 for the error function nonlinearity:

$$\bar{r} = \frac{r_{max}}{2} \text{erfc}\left( \frac{y_0}{\sigma\sqrt{2}} \right)$$ \hspace{1cm} (10)

$$C = \frac{r_{max}}{2} \frac{\sigma\delta}{\sqrt{\pi}} \exp\left( -\frac{y_0^2\delta^2}{2\sigma^2} \right)$$ \hspace{1cm} (11)

where $\delta = [\varepsilon/\sigma]^{-1} + 1^{1/2}$. Because these equations only provide two constraints and the nonlinearity has three parameters, additional information is required to identify the system. One approach is to obtain additional constraints by estimating...
higher moments between the input and output of the system, which is discussed below. Another possibility is to begin by estimating \( r_{max} \). For example, \( r_{max} \) can be taken as the reciprocal of the minimum inter-spike interval in the data record, which is the method used in this paper. Alternatively, the maximum output rate may be determined by the response to optimal stimuli from other experiments. In the case of 2AFC psychophysical experiments, \( r_{max} \) can be defined as 1. Thus, if we know \( r_{max} \) and measure \( r \) and \( C \), we can solve for \( \gamma_0 \) and \( \varepsilon \) analytically.

**Naka-Rushton Nonlinearity**

A final example of a sigmoidal function commonly used in models is the Naka-Rushton (NR) equation,

\[
g(y) = \frac{r_{max}y^n}{y^n + c^n}
\]

with parameters \( r_{max}, c, \) and \( n \). Unfortunately, the integrals in Equations 2 and 4 do not have nice analytic expressions, and we can still solve numerically the system of integral equations,

\[
r = \frac{r_{max}}{\sqrt{2\pi}\sigma^2} \int \frac{y^n}{y^n + c^n} \exp\left(\frac{-y^2}{2\sigma^2}\right) dy
\]

\[
C = \frac{r_{max}nc^n\sigma}{\sqrt{2\pi}} \int \frac{y^{n-1}}{(y^n + c^n)^2} \exp\left(\frac{-y^2}{2\sigma^2}\right) dy
\]

for \( c \) and \( n \) provided that \( r_{max} \) is estimated from some additional information as above.

**Results**

To demonstrate the ability of the moment method to reconstruct the nonlinearity, we tested the model with two simulated experiments. First, we modeled the response of a human subject in a 2AFC psychophysical task. Second, we modeled the receptive field of a simple-cell in primary visual cortex. In each case, we calculated the mean response rate and correlation magnitude, using the method described in detail in Appendix B. To evaluate the performance of the method, the parameters of the simulated nonlinear system were compared with these estimated values. We also compared the performance of the proposed technique with variants of the linear reconstruction method. Anzai et al. (1999) used a straightforward implementation of the linear reconstruction (or LR) method, whereas Chichilnisky (2001) implemented a cross-validated version of the technique (or CLR). Details about the implementations of these algorithms are described in Appendix C. To anticipate the results, we find that the moment method clearly outperforms the LR and CLR techniques.

**Psychophysical Task**

We simulated the study of Ahumada & Beard (1998) who measured classification images in a Vernier acuity task. In our simulations, we assume a spatial kernel (the vector \( w \)) similar to those obtained in the actual measurements (Figure 1a) and an error function nonlinearity with \( r_{max} = 1, \gamma_0 = 0.5, \) and \( \varepsilon = 1.0 \). In each trial of the experiment one of two possible stimuli, \( x_0 \) or \( x_1 \), is presented in the presence of additive white Gaussian noise with \( \sigma = 1 \). The subject is asked to identify the stimulus. We code the subject’s response as \( r = 0 \) for the \( x_0 \) choice and \( r = 1 \) for the \( x_1 \) choice. The probability that the subject will select \( x_1 \) is modeled as \( Pr(r = 1 \mid x_1) = g(w'x_1) \). If we pool all the trials in which only \( x_1 \) was presented, we obtain \( Pr(r = 1 \mid x_1) = g(w'x_1 + n') = g(\theta_0 + w'n) \) (where \( n \) represents the noise in the stimulus). Similarly, for the trials in which only \( x_0 \) was presented, we obtain \( Pr(r = 1 \mid x_0) = g(w'x_0 + n) = g(\theta_0 + w'n) \). The “classification images” obtained by cross-correlation between response and stimuli in these two cases should be the same and proportional to \( w \).

Similarly, the nonlinearities should be identical up to a translation. In fact, these constraints could be used as a test for the validity of the model in a 2AFC task.

We ran simulations that contained between 300 and 100,000 trials per stimulus. We selected one of the stimuli and calculated the linear kernel (or “classification image”) (Figure 1b and 1c) and the estimated error function nonlinearity (Figure 2). The nonlinearity calculations show that the proposed method converges faster than the LR/CLR methods (the result from the CLR method is not shown, but it was no better than the LR method). Our technique gives a reasonable approximation to the nonlinearity after only 500 trials, whereas the LR method yields a gross underestimate (Figure 2a). After 2400 trials, the moment method has determined the nonlinearity almost perfectly, but the LR method still underestimates (Figure 2b). After 75,000 trials, both methods have converged to the correct solution (Figure 2c).

To study the convergence of the algorithm, we simulated 60 runs and estimated the average relative error of the parameters for increasing lengths of the data record, ranging from 300 to 100,000 trials. A summary of the average relative error in \( \gamma_0 \) and \( \varepsilon \) vs. length of the data record is shown in Figure 2d. After 2500 trials, the error is down to 10 percent, and it converges toward zero as the number of trials is increased further.

**Simple Cell**

The second simulation was that of a simple cell in visual cortex. To mimic a cell’s response, we used a linear kernel obtained from cross-correlation analysis of actual experimental data. The kernel was discretized in 45 time slices with \( \Delta T = 2 \) ms and by a \( 40 \times 40 \) grid in space (see Figure 3a). In each 2-ms interval, the probability of a spike was given by Equation 1, and the nonlinearity was a power law with parameters \( A = 0.01 \) and \( B = 2 \). The input was binary noise taking the values 1 and –1, giving a standard deviation of \( \sigma = 1 \).

We simulated the response of the simple cell to noise stimuli with experimental time ranging from 1 minute to 8 hours. For each case, we estimated the kernel (Figure 3b and 3c) and the nonlinearity almost perfectly, but the LR method still underestimates (Figure 4). These simulations show the strength of the moment method with large kernels, such as the \( 40 \times 40 \times 45 \) kernel simulated here. Unlike the LR/CLR methods, our technique closely approximates the simulated nonlineari-
ty after only 15 minutes, or less than 5000 spikes (Figure 4a).

Even the CLR method required 8 simulated hours (more than
140,000 spikes) to converge. The bias in the LR method is large
and is still substantial in the 8-hour simulation.

The relatively poor performance of the CLR method can be
explained by the noise in the estimate of a large kernel. If the
estimate of $w$ were extremely noisy, the linear prediction
$y = w^T x$ would be mostly noise, and the CLR method would
give a flat line at the mean rate. Figure 5b illustrates that the
CLR results lie between the mean rate and the true nonlinearity
after 15 minutes of data collection, as would be expected from a
noisy estimate of $y$. As the simulation length increases, the
results approach the correct nonlinearity. A more detailed dis-
cussion of the noise in the CLR methods, as well as a discussion
on the bias in the LR method, can be found in Appendix C.

Figure 4d shows the rate of convergence of the estimated
nonlinearity as a function of simulated time between 1 minute
and 8 hours. For each data point of 1 hour or less, we computed
12 simulations and calculated the average relative error in the
nonlinearity parameters. For longer trials, the relative error esti-
mates were based on four simulations of 2 hours, two simula-
tions of 4 hours, and one simulation of 8 hours. After only
10 minutes of simulation (around 3000 spikes), the estimated
parameters of the power law are within 10 percent of the
actual parameters.

To obtain the above results, we assumed we knew the
nonlinearity was a power law. Normally, the investigator does
not necessarily know in advance the shape of the nonlineari-
ty. A natural question is what would happen if we were to fit
other nonlinearities to data generated by a power law func-
tion. Figure 5a shows results of the moment method if we
assumed a power law, error function, or Naka-Rushton non-
linearity. The different estimates roughly agree for small val-
ues of $y$ but diverge as $y$ increases. Thus, confidence in any
extrapolation to large $y$ values requires precise knowledge of
the shape of the nonlinearity. Because the LR and CLR meth-
ods do not make any assumptions about the shape of the non-
linearity, they could be used to determine the proper class of
nonlinearities to use in the moment method.

The moment method is a parametric technique that
assumes a particular class of nonlinearities. On the other hand,
the LR and CLR techniques are nonparametric algorithms. A
natural question is: would the performance of the two tech-
niques be similar if we added information about the class of
relevant nonlinearities to the LR/CLR methods. This could be
done by fitting a parametric nonlinear function to the scatter-
plot of the linear prediction versus the actual response in the
LR/CLR algorithms. In the simulation of the psychophysics
experiment (Figure 2), we fit an error function, and for the
simulation of the simple-cell receptive field (Figure 4), we fit a
power-law. Figure 6 shows the maximum relative error of the
parameters obtained from these fits as well as the parameters
from the moment method. In all cases, the moment method is
dramatically superior to the LR/CLR methods. The better per-
formance of the LR method than the CLR method in Figure 6a
is understandable because removing bias can sometimes
decrease accuracy. The erratic behavior of the LR results in
Figure 6b is due its deviation from a power-law shape (as seen
in Figure 4a and b).

**Discussion**

The moment method is a fast and accurate method to
identify a linear-nonlinear system. In this method, we estimate
the linear filter using cross-correlation, and identify the output
nonlinearity by matching the mean response and cross-correla-
tion amplitude to the data. The method is fast because it does
not require the calculation of a linear prediction to estimate
the static nonlinearity. Simulations demonstrate that the rate
of convergence of the moment technique is faster than the LR
and CLR methods. The advantage of the moment method is

---

**Figure 1. The kernel used for the psychophysical task.**

*a.* The kernel used in the simulation. Red regions indicate areas where increases in luminance would result in an increase in the probability of reporting the $x_1$ stimulus. Blue regions represent areas where increases in luminance would result in a decrease of the probability of reporting the $x_1$ stimulus. 
*b-c.* The estimate of the kernel after 10,000 (b) and 100,000 (c) trials.
particularly evident when the kernels are large, as in the simulation of the simple cell (Figure 4). Here, the moment method obtains a reasonable estimate of the nonlinearity after 15 minutes, whereas the LR/CLR estimates are inaccurate for this amount of data. Another benefit of the technique is that estimates can be updated recursively as new data arrives. This provides a means to stop an experiment when a sufficient signal-to-noise ratio is achieved. We note that both the moment method and the LR/CLR methods are dependent on the assumption of a linear-nonlinear system as described by Equation 1. Thus, these methods will give accurate results only to the extent that a system is well represented by this lumped phenomenological model.

As mentioned above, one advantage of the LR/CLR methods is that no assumptions are made about the shape of the nonlinearity. This observation suggests using the LR/CLR methods initially when studying a new system. These experiments will be time consuming, as large amounts of data are required for accurate estimates of \( g(\cdot) \). However, once a family of parametric functions is found to fit the measured nonlinearities, the more efficient moment method can be used to conduct further experiments.

Several nonlinearities of interest are parameterized by more than two variables. In these cases, additional conditions must be provided to fully determine \( g(\cdot) \). One possibility is to obtain independent information about the remaining parame-
In the present study, we calculate two moments to estimate a two-parameter nonlinearity. We are currently investigating the possibility of calculating higher order moments to constrain the additional parameters.

### Appendix A. Derivation of Equations for Nonlinearity

In this appendix, we derive the equations for the mean response $\bar{r}$ and correlation $E.xr$ for the linear-nonlinear system in response to a discrete approximation of Gaussian white noise. In this derivation, we ignore the distinction between space and time and assume that the probability of a response is given by

$$\rho_x(x) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\frac{||x||^2}{2\sigma^2}\right),$$

(A2)

We derive Equation 2 for the mean response $\bar{r}$:

$$\bar{r} = \frac{1}{(2\pi\sigma^2)^{N/2}} \int_{\mathbb{R}^N} g(w^T x) \exp\left(-\frac{||x||^2}{2\sigma^2}\right) dx$$

$$= \frac{1}{(2\pi\sigma^2)^{N/2}} \int_{\mathbb{R}^N} g(u_i) \exp\left(-\frac{||u||^2}{2\sigma^2}\right) du$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} g(y) \exp\left(-\frac{y^2}{2\sigma^2}\right) dy.$$  

(A3)

In the derivation, we let $u = Ox$, where $O$ is any orthogonal matrix whose first row is $w^T$. (Recall that $w$ is a unit vector.) Then $w^T x$ is the first component of $u$, which we replaced by $y$.

To derive an expression for the input-output correlation

$$E.xr = \frac{1}{(2\pi\sigma^2)^{N/2}} \int_{\mathbb{R}^N} x g(w^T x) \exp\left(-\frac{||x||^2}{2\sigma^2}\right) dx$$

we first look at

$$B = E.(Ox y) = \frac{1}{(2\pi\sigma^2)^{N/2}} \int_{\mathbb{R}^N} u g(u_i) \exp\left(-\frac{||u||^2}{2\sigma^2}\right) du$$

(A4)

we first look at

$$B_j = 0$$

for $j \neq 1$ because $B_j$ is an odd integral in this case. The first component of $B$ is

$$B_1 = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} g(y) \exp\left(-\frac{y^2}{2\sigma^2}\right) dy$$

$$= \frac{\sigma}{\sqrt{2\pi}} \int g'(y) \exp\left(-\frac{y^2}{2\sigma^2}\right) dy,$$

where we integrated by parts to obtain the last expression. Thus,

$$B = \frac{\sigma}{\sqrt{2\pi}} \int g'(y) \exp\left(-\frac{y^2}{2\sigma^2}\right) dy e_1.$$

(A7)

where $e_1$ is the first unit vector. The input-output correlation is then

$$E.xr = O^T B = C.w$$,

(A8)

where

$$C = \frac{\sigma}{\sqrt{2\pi}} \int g'(y) \exp\left(-\frac{y^2}{2\sigma^2}\right) dy.$$

(A9)

This result is known as Bussgang's Theorem (Bussgang, 1952).

### Half-Rectifier With Threshold Nonlinearity

If $g(y) = A[y - y_t]^+$, then

$$C = \frac{A\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{y^2}{2\sigma^2}\right) dy$$

$$= \frac{A\sigma^2}{2} \text{erfc}\left(\frac{y_t}{\sqrt{2\sigma^2}}\right).$$

(A10)
Figure 4. Nonlinearity calculations from the simple-cell simulation. a. Estimated nonlinearity after 15 minutes of simulated time. The solid green line is the power-law estimate of the nonlinearity and the dashed blue line is the nonlinearity used in the simulation. The error bars of the LR (red) and CLR (black) represent one standard error and were calculated from the second step of the reconstruction. Parameters from the estimated nonlinearity: $A = 0.00959$ and $\beta = 2.08$. b. Estimated nonlinearity after 1 hour of simulated time. Estimated parameters: $A = 0.00985$ and $\beta = 2.02$. c. Estimated nonlinearity after 8 hours of simulated time. Estimated parameters: $A = 0.00995$ and $\beta = 2.002$. d. Rate of convergence of the moment method to the simulated parameters. Relative errors are less than 10% after 10 minutes of simulated time. In panels a-c, the y axis is in units of spikes per ms.

Figure 5. a. Estimated error function (red line), power law (green line), and Naka-Rushton (cyan line) nonlinearities to one simulated hour of data generated by a power law function (dashed blue line). The y axis is in units of spikes per ms. b. Consequences of a noisy estimate of $w$ on the shape of the nonlinearity obtained by the CLR method. The dotted line represents the mean output rate. As expected, the estimate lies between this line and the simulated nonlinearity. Data are from 15 minutes of simulated time.
Figure 6. Rates of convergence of the methods to estimate the nonlinearity parameters. Relative errors in the moment method (green line), LR method (red line), and CLR method (black line) are plotted versus simulation length. a. Relative error of fitted error function parameters in the psychophysical task simulation. The maximum of the relative error in $y_c$ and the relative error in $\varepsilon$ is shown. Relative errors are less than 10% after 2500 trials for the moment method and after 40,000 to 50,000 trials for the LR and CLR methods. Results are from the same simulations used for Figure 2d. b. Relative error of fitted power law parameters in the simple cell simulation. The maximum of the relative error in $A$ and the relative error in $\beta$ is shown. Relative errors are less than 10% after 10 minutes of simulated time for the moment method and after 8 hours for the CLR method. The LR results were not fit well by a power law after less than 2 hours of simulated time; thus, the relative error measurements for those simulations have little meaning. Even after 8 hours, the relative error for the LR method was around 30%. Results are from the same simulations used for Figure 4d.

Error Function Nonlinearity

Let the nonlinearity be

$$g(y) = \frac{r_{\text{max}}}{2} \left[ 1 + \text{erf} \left( \frac{y - y_0}{\varepsilon \sqrt{2}} \right) \right]$$

(A14)

where $\text{erf}(x)$ is the error function $\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2) dt$. The derivative of $g(y)$ is

$$g'(y) = -\frac{r_{\text{max}}}{\sqrt{2} \pi \varepsilon^2} \exp \left( -\frac{(y - y_0)^2}{2 \varepsilon^2} \right) .$$

Then, the mean response is

$$r = \frac{1}{\sqrt{2 \pi \sigma^2}} \int_{-\infty}^{\infty} \left[ 1 + \text{erf} \left( \frac{y - y_0}{\varepsilon \sqrt{2}} \right) \right] \exp \left( -\frac{y^2}{2 \sigma^2} \right) dy$$

(A15)

$$= \frac{1}{\sqrt{2 \pi \sigma^2}} \int_{-\infty}^{\infty} \exp \left( -\frac{(y - y_0)^2}{2 \varepsilon^2} \right) \text{erfc} \left( \frac{y}{\varepsilon \sqrt{2}} \right) dy$$

$$= \frac{r_{\text{max}}}{\sqrt{2 \pi \sigma^2}} \int_{-\infty}^{\infty} \exp \left( -\frac{(y - y_0)^2}{2 \varepsilon^2} + \frac{y^2}{2 \sigma^2} \right) dy$$

$$= \frac{r_{\text{max}}}{\sqrt{2 \pi \sigma^2}} \text{erfc} \left( \frac{y_0 \delta}{\sqrt{2 \sigma^2}} \right)$$

where $\text{erfc}(x) = 1 - \text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} \exp(-t^2) dt$ and $\delta = \sqrt{[\varepsilon/\sigma^2] + 1}^{-1/2}$. We integrated by parts in the second step. To obtain the last step, we computed a change of variables with one of the variables parallel to the line $u = (y + y_0)/(\sigma \sqrt{2})$ and completed the square in the resulting integrand. We have thus obtained Equation 10.

The magnitude of the input-output correlation is obtained by a completion of the square:

\[ \int_{-\infty}^{\infty} \exp(-t^2) dt = \sqrt{\pi} \]

which is Equation 6. Also,

\[ r = \frac{A}{\sqrt{2 \pi \sigma^2}} \int_{-\infty}^{\infty} \exp \left( -\frac{y^2}{2 \sigma^2} \right) dy \]

(A12)

\[ = \frac{A}{\sqrt{\pi}} \sigma^2 e^{2 - 2 \beta \gamma} \int_0^{\infty} \exp \left( -u^2 \right) du \]

\[ = \frac{A}{\sqrt{\pi}} \sigma^2 e^{2 - 2 \beta \gamma} \Gamma(\beta + 1) / 2 \]

which is Equation 7. Also,

\[ C = \frac{A \beta \sigma}{\sqrt{2 \pi}} \int_0^{\infty} \exp(-y^2 / 2 \sigma) dy \]

(A13)

\[ = \frac{A \beta}{\sqrt{\pi}} \sigma^{\beta + 1 / 2} e^{2 - 2 \beta \gamma} \int_0^{\infty} u^{\beta - 1 / 2} e^{-u} du \]

\[ = \frac{A \beta}{\sqrt{\pi}} \sigma^{\beta + 1 / 2} e^{2 - 2 \beta \gamma} \Gamma(\beta / 2) \]

which is Equation 8.

Power Law Nonlinearity

If $g(y) = A y^\beta$, then

\[ r = \frac{A}{\sqrt{2 \pi \sigma^2}} \int_{-\infty}^{\infty} y^\beta \exp \left( -\frac{y^2}{2 \sigma^2} \right) dy \]

(A11)

\[ = \frac{A \sigma^{\beta + 1}}{\sqrt{\pi}} \int_0^{\infty} u^{\beta - 1 / 2} e^{-u} du \]

\[ = \frac{A \sigma^{\beta + 1 / 2}}{\sqrt{\pi}} \Gamma(\beta / 2) \]

\[ \text{Error Function Nonlinearity} \]

\[ g(y) = r_{\text{max}} \left[ 1 + \text{erf} \left( \frac{y - y_0}{\varepsilon \sqrt{2}} \right) \right] \]

(A14)
Because the argument of the nonlinearity \( g(x) \) is relatively small, the central limit theorem permits the approximation to Gaussian white noise as described above will have a deviation, \( \sigma \), equal to

\[
C = \frac{\sigma}{\sqrt{2\pi} \cdot \sqrt{2\pi \sigma^2}} \int \exp\left( -\frac{(y - y_0)^2}{2\sigma^2} \right) \exp\left( -\frac{y^2}{2\sigma^2} \right) dy
\]

\( = r_{\text{max}} \frac{\sigma}{\sqrt{2\pi}} \exp\left( -\frac{y_0^2}{2\sigma^2} \right). \quad (A16)

**Non-Gaussian Input**

One challenge in using the reverse-correlation technique is getting enough data for a reasonable signal-to-noise ratio. Typically, one must run long experiments in order to obtain good results. This problem is exacerbated with physiological experiments because cells do not respond well to a white noise stimulus. Thus, one would like to develop ways to increase the output rate.

Let \( M \) denote the maximum stimulus magnitude that could be produced by an input device, ie, each component of the stimulus must be in the interval \([-M, M]\). In that case, a preferable input would be one where each component of the stimulus was a binary random variable that took on the values \( \pm M \). Typically, one must run long experiments in order to obtain good results. This problem is exacerbated with physiological experiments because cells do not respond well to a white noise stimulus. Thus, one would like to develop ways to increase the output rate.

Then \( \sigma \) is the standard deviation of the stimulus that increases the stimulus standard deviation.

Thus, one would like to develop ways to increase the output rate. If \( g(x) \) is equal to 0, the output rate increases with the standard deviation \( \sigma \) of the input [to see this, integrate Equation 2 by parts]. Therefore, one would like to maximize the standard deviation of the input. However, physical devices, like computer monitors, have limited dynamic ranges. The discrete approximation to Gaussian white noise as described above will have a relatively small \( \sigma \) given restraints on the magnitude of the stimulus. We thus extended the theory to more general forms of noise stimuli that increase the stimulus standard deviation. We used Bayes’ Theorem to complete the third step and Equations 1 and A2 in the fourth step. \( \rho_{x \rightarrow y}(x) \) is the probability density function of \( x \) conditioned on \( r = 1 \). The mean output rate is simply \( \bar{r} = \int \rho \). The correlation magnitude is thus

\[
C = \left\| E[\mathbf{x}r] \right\| = \| \mathbf{p} \| \bar{r}. \quad (B3)
\]

There is a subtlety regarding how one should calculate \( \| \mathbf{p} \| = \mathbf{p}^T \mathbf{p} \) to reduce the bias in the estimate of \( C \). If \( \mathbf{p} \) is a noisy estimate of \( \mathbf{p} \), then \( \| \mathbf{p} \| \) overestimates the magnitude of \( \| \mathbf{p} \| \). Thus, a naive calculation of \( C \) will overestimate the input-output correlation.

Assume \( \tilde{\mathbf{p}} \) is an unbiased estimator of \( \mathbf{p} \) so that \( E[\tilde{\mathbf{p}}] = \mathbf{p} \). Let \( p_j \) be the jth component of \( \mathbf{p} \). If \( \sigma_p \) is the standard deviation of \( \tilde{\mathbf{p}}_j \), then

\[
p_j^2 = E\{\tilde{p}_j^2\} - \sigma_p^2 \quad (B4)
\]

and

\[
\mathbf{p}^T \mathbf{p} = E\{\tilde{\mathbf{p}}^T \tilde{\mathbf{p}}\} - (\sigma_p)^T \sigma_p. \quad (B5)
\]

Thus, an unbiased estimate of \( \| \mathbf{p} \| \) is

\[
\| \mathbf{p} \|^2 = \tilde{\mathbf{p}}^T \tilde{\mathbf{p}} - (\sigma_p)^T \sigma_p. \quad (B6)
\]

To calculate \( \sigma_p \), we divide the results into multiple trials and calculate an approximation to \( \mathbf{p} \) for each trial. We let \( \tilde{\mathbf{p}} \) be the average of these values and \( \sigma_p \) be the sample standard deviation of the mean. We then use B6 and B3 to calculate the input-output correlation.

The bias with the naive estimate of the input-output correlation increases with the size of the kernel. The magnitude \( \| \mathbf{p} \| \) is fixed for a given nonlinearity, but \( \| \mathbf{p} \| \) increases with the number of components. Thus, the bias in the naive estimate of \( C \) was much larger with the \( 40 \times 40 \times 45 \) kernel of the simple cell than with the \( 32 \times 32 \) kernel of the psychophysical task.

Calculations with only small amounts of data can be so noisy that \( (\sigma_p)^T \sigma_p > \tilde{\mathbf{p}}^T \tilde{\mathbf{p}} \). In that case, \( \| \mathbf{p} \| \) is the square root of a negative number, and we cannot compute an estimate of the nonlinearity. Thus, in Figure 2d we could not include the results for a few simulations with 500 or fewer trials. Similarly, we could not include one simulation that was a minute long in Figure 4d.
Appendix C: Linear Prediction Method

For a benchmark by which to evaluate the moment method, we also calculated the static nonlinearity with a linear reconstruction (LR) method similar to that employed by Anzai et al (1999) and Chichilnisky (2001). This two-step method involves first calculating the linear kernel \( w \) via reverse correlation and then calculating the average response as a function of the linear prediction \( y = w^T x \). In the latter calculation, we discretized \( y \) into intervals, using narrower intervals around zero and wider intervals away from zero, because \( y \) will have a normal distribution centered around zero. For each interval, we calculated the mean and standard error of the response if the interval contained three or more data points.

We present results from both the LR method and a cross-validated implementation of the linear-prediction method (CLR). In the LR method, we used all the data for both parts of the calculation. In the CLR method, we divided the data into \( N \) parts, where \( N = 200 \). We calculated \( N-1 \) versions of the linear kernel, where each version used all the data except two consecutive parts. Thus, each version of the kernel would be based on approximately 99% of the data. We then used all the data to calculate the average response as a function of the linear prediction. The key of the CLR method was to switch kernels used for the linear prediction \( y = w^T x \), so that the same data was never used simultaneously for both the linear kernel estimate \( w \) and the corresponding stimulus \( x \). Thus, if the CLR method removes the bias in \( y = w^T x \) created from covariance between the stimulus \( x \) and the noisy estimate \( w \).

The bias in \( y = w^T x \) is similar to the bias in \( \|p\|_2 = p^T p \) discussed above. As with \( \|p\|_2 \), the bias in \( w^T x \) increases with the size of the kernel. Thus, for small kernels, the CLR method may not give better results than the LR method because sometimes removing bias can increase the error. The bias with large kernels, on the other hand, may greatly skew the results of the LR method. In this case, the CLR method will be significantly more accurate. For the small \( 32 \times 32 \) kernel of the psychophysical simulations, the bias in the LR method did give an underestimate of the nonlinearity (Figure 2a,b), but the bias was small. As shown in Figure 6a, the LR method outperformed the CLR method in this case. However, the large \( 40 \times 40 \times 45 \) kernel of the simple-cell simulations produced significant bias in the LR method even after 8 hours (Figure 4). In this case, the CLR method was necessary to give accurate results.

Although the CLR method removes the bias caused by covariance between the stimulus and the kernel estimate, it is still sensitive to noise in the estimate of \( w \), as discussed in connection with Figure 5b. When \( w \) is high dimensional, small noise in each component of \( w \) can build up to make the unbiased estimate of \( y = w^T x \) very noisy, as observed in the simple-cell simulations.

Acknowledgments

We thank E. J. Chichilnisky and J. D. Victor for helpful comments and criticisms on an earlier version of this paper. This work was supported by an NSF Mathematical Sciences Postdoctoral Research Fellowship (D.Q.N) and by NSF-IBN-9720305 and NIH-EY-12816 (D.L.R).

References


