

## Adaptive Spatiotemporal Receptive Field Estimation in the Visual Pathway

Garrett B. Stanley

*gstanley@deas.harvard.edu*

*Division of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, U.S.A.*

The encoding properties of the visual pathway are under constant control from mechanisms of adaptation and systems-level plasticity. In all but the most artificial experimental conditions, these mechanisms serve to continuously modulate the spatial and temporal receptive field (RF) dynamics. Conventional reverse-correlation techniques designed to capture spatiotemporal RF properties assume invariant stimulus-response relationships over experimental trials and are thus limited in their applicability to more natural experimental conditions. Presented here is an approach to tracking time-varying encoding dynamics in the early visual pathway based on adaptive estimation of the spatiotemporal RF in the time domain. Simulations and experimental data from the lateral geniculate nucleus reveal that subtle features of encoding properties can be captured by the adaptive approach that would otherwise be undetected. Capturing the role of dynamically varying encoding mechanisms is vital to our understanding of vision on the natural setting, where there is absence of a true steady state.

### 1 Introduction

In the early visual pathway, each neuron encodes information about a restricted region of visual space that is generally referred to as the receptive field (RF) of the cell. The dominant paradigm over the past several decades is that the RF extent and the corresponding temporal encoding properties of the neuron are primarily a function of bottom-up processing, and are thus invariant properties of the cell. Overwhelming evidence contradicts this static construct, pointing toward more complex time-varying encoding mechanisms that result from a variety of dynamic sources. Adaptation and plasticity exist on a number of different timescales ranging from milliseconds to hours; these mechanisms have been studied for some time, but the role in the encoding process is not yet known. In this article, a new approach to RF estimation is presented that is critical in precisely quantifying effects of systems-level plasticity or adaptation and the role they play in encoding information about the visual world. An adaptive implementation

is developed that characterizes the functional nature of the time-varying properties that have been explored experimentally but have not yet been well quantified in the context of neuronal encoding.

The linear correlation structure between controlled sensory stimuli and the corresponding recorded neural activity has been shown to reveal a significant amount of information about the underlying functional properties of the neuron. Reverse-correlation techniques, which refer to the cross-correlation between stimulus and response for white noise stimuli, have been implemented in a number of systems to characterize the linear component of the stimulus-response relationship of the cells. In the visual pathway, spatiotemporal white noise stimuli have been used extensively to characterize the dynamics of neurons in the retina, lateral geniculate nucleus (LGN), and primary visual cortex using the reverse-correlation technique (Jones & Palmer, 1987; Reid, Soodak, & Shapley, 1991; Reid, Victor, & Shapley, 1997; Mao, MacLeish, & Victor, 1998). Additionally, in the auditory system, researchers have long since used reverse-correlation techniques for characterizing auditory neurons in the periphery (DeBoer & Kuyper, 1968), and others have recently used reverse-correlation techniques to extract sound-feature selectivity of neurons in the auditory cortex (deCharms, Blake, & Merzenich, 1998). These techniques have been instrumental in developing our understanding of systems-level processing in the early sensory pathways. One of the major assumptions of the reverse correlation technique is that the dynamics of the underlying functional mechanisms in the visual pathway are time invariant, resulting in spatiotemporal RF properties that remain unchanged with time. As early as the retina, however, adaptation mechanisms act on a continuum of timescales to adjust encoding dynamics in response to changes in the visual environment (Enroth-Cugell & Shapley, 1973; Shapley & Victor, 1978). Adaptation mechanisms have also been identified in the LGN (Ohzawa, Sclar, & Freeman, 1985) and cortical area V1 (Albrecht, Farrar, & Hamilton, 1984), and models have been suggested in which population activity acts in a divisive manner to regulate the cellular response properties (Heeger, 1992). Furthermore, several recent studies have shown that the spatial and temporal characteristics of visual neurons can vary drastically over a variety of different timescales (Pei, Vidyasagar, Volgushev, & Creutzfeld, 1994; Gilbert, Das, Ito, Kapadia, & Westheimer, 1996; Mukherjee & Kaplan, 1995; Przybyszewski, Gaska, Foote, & Pollen, 2000), through neuromodulatory influences from other brain regions. Neurons throughout the visual pathway have been shown to exhibit temporal and spatial properties that evolve over a continuum of timescales, bringing into question the reliability of traditional reverse-correlation techniques applied over even short experimental trials. In this article, an adaptive approach for the estimation of spatiotemporal RF properties from single trials is developed specifically to address time-varying dynamics in the visual pathway, and the utility is demonstrated in simulation of geniculate and cortical dynamics, as well as experimental data from the cat LGN. Portions

of these results have been presented elsewhere in abstract form (Stanley, 2000b).

## 2 Receptive Field Estimation

---

The foundation for much of modern systems neurophysiology is the characterization of the relationship between sensory inputs and the corresponding evoked neural activity. This relationship is often referred to as the RF and in general encompasses temporal as well as spatial characteristics of the underlying sensory map. Temporal estimation is first developed here, but is easily extended to complete spatiotemporal maps. Suppose the spike train of a neuron is denoted as  $\rho(t) = \sum_j \delta(t - \tau_j)$ , with spike event times  $\tau_0, \tau_1, \tau_2, \dots$  on the interval  $(0, t_f]$ , where  $\delta$  is the Dirac delta function. The fundamental task lies in the characterization of the relationship between a temporally continuous stimulus and the discrete process of the firing activity of the neuron. One approach is to relate the continuous input to the firing “rate” of the neuron, allowing the quantification of the relationship between the input and its modulatory effects on the rate of action potentials generated.

Let the firing rate of the neuron be denoted  $r[k]$ , which represents the number of events occurring in the interval  $((k-1)\Delta t, k\Delta t]$  normalized by the interval width,  $\Delta t$ . The rate is also often interpreted as the time-dependent rate of an inhomogeneous Poisson process, which obviously results in discrete stochastic event times (Ringach, Sapiro, & Shapley, 1997). Although the transformation to rate does simplify the relationship a great deal, the remaining relationship between the stimulus, which takes on values both positive and negative relative to the mean level, and the strictly nonnegative firing rate is nontrivial. Linear models are insufficient to capture such a transformation, and thus generally must be described through a higher-order Wiener kernel expansion. An alternate functional model to describe the relationship between the stimulus and the firing rate of a neuron is the Wiener system, which incorporates a linear (L) system followed by a static nonlinearity (N), and is therefore often also referred to as an LN system (Movshon, Thompson, & Tolhurst, 1978; Tolhurst & Dean, 1990; Ringach et al., 1997; Chichilnisky, 2001), as shown in Figure 1. The output of the linear element is expressed as a convolution of the temporal stimulus,  $s$ , with the

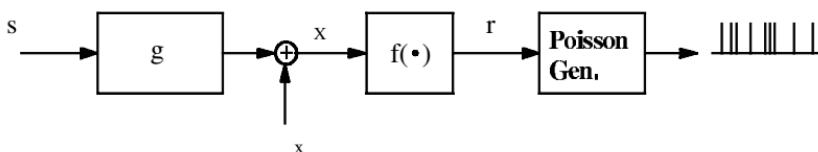


Figure 1: Wiener system.

first-order kernel  $g$ , plus an offset  $\mu_x$ ,

$$x[k] = \sum_{m=0}^{L-1} g[m]s[k-m] + \mu_x,$$

where  $L\Delta t$  is the filter length, which can be interpreted as the temporal integration window of the cell. For white noise inputs, since  $x$  is the linear combination of independent random variables, the process  $x$  is approximately gaussian. The firing rate of the neuron is then a nonlinear function of the intermediate signal  $x$ , so that  $r[k] = f(x[k])$ . The inherent rectifying properties of neurons can often be described through a simple half-wave rectification function,  $\lfloor \cdot \rfloor$ , resulting in the output  $r[k] = \lfloor x[k] \rfloor$ , which is then driving an inhomogeneous Poisson process. In contrast to the complex nature of a higher-order Wiener kernel representation, the Wiener system provides a relatively simple means for describing the inherent nonlinearity in neural encoding. The basic characteristics of the Wiener system have been discussed in a variety of studies focused on a number of systems. Several studies have provided algorithms for the identification of Wiener cascades (Hunter & Korenberg, 1986; Korenberg & Hunter, 1999), and the convergence properties of the estimators have also been well characterized (Greblicki, 1992, 1994, 1997). Although these developments were targeted at more general physiological systems, the utility of these techniques in neural systems has been recently demonstrated (Paulin, 1993; Dan, Atick, & Reid, 1996; Ringach et al., 1997; Smirnakis, Berry, Warland, Bialek, & Meister, 1997; Chichilnisky, 2001). In the work presented here, an adaptive approach to Wiener system estimation is used to capture time-varying characteristics in the dynamic relationship between visual stimulus and neuronal response.

**2.1 Estimation of Time-Invariant Properties.** Before the adaptive approach is developed, the framework for the general estimation problem must first be discussed. Let the parameter vector be defined as  $\theta \triangleq [g[0] \ g[1] \ \dots \ g[L-1]]^T \in \mathbb{R}^{L \times 1}$  to represent the first-order kernel, where  $T$  denotes transpose. The response can then be written  $r[k] = \lfloor x[k] \rfloor$ , where  $x[k] = \theta^T \varphi[k] + \mu_x$  and  $\varphi[k] \triangleq [s[k] \ s[k-1] \ \dots \ s[k-L+1]]^T$  is the time history of the stimulus. As is well known and shown for reference in section A.1, the parameter vector for a first-order kernel can be estimated from the cross-covariance between the stimulus and the output of the linear element, normalized by the autocovariance structure of the stimulus, or  $\hat{\theta} = \Phi_{ss}^{-1} \phi_{sx}$ , where  $\Phi_{ss} \in \mathbb{R}^{L \times L}$  is the Toeplitz matrix of the autocovariance of the stimulus  $s$ ,  $\phi_{sx} \in \mathbb{R}^{L \times 1}$  is the cross-covariance vector between the stimulus and the output of the linear block,  $x$ , and  $\hat{\cdot}$  denotes estimate. As discussed in section A.2, for gaussian inputs to the static nonlinearity, the only effect is a scaling of the cross-covariance, and the estimate of the parameter vector

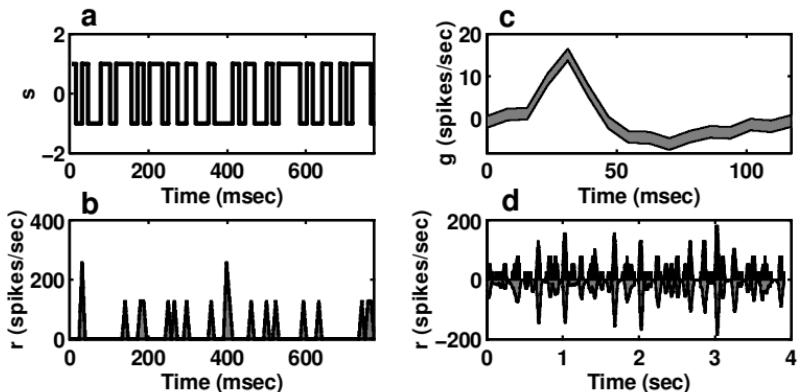


Figure 2: First-order kernel estimation. The first-order kernel was computed from (a) the temporal m-sequence and (b) the corresponding firing rate (in 7.7 ms bins) of X cell in the LGN. (c) The temporal kernel at the center of the receptive field; band represents  $\pm 2SD$  around estimate (see equation A.1). The actual (positive) and predicted (negative) firing rates (in 40 msec bins) from the full spatiotemporal RF model. Approximately 4 minutes of data at a sampling rate of 128 Hz.

can be expressed as a function of the stimulus,  $s$ , and response,  $r$ ,

$$\hat{\theta} = \Phi_{ss}^{-1} \phi_{sx} = \frac{1}{C} \Phi_{ss}^{-1} \phi_{sr}, \quad (2.1)$$

where  $\frac{1}{C} \in \mathbb{R}$  is a constant of proportionality. As shown in section A.2, for half-wave rectification with gaussian inputs with mean much less than the standard deviation, this scaling will be approximately 2, which is consistent with the experimental results presented here. Figure 2 shows the estimate of a low-pass biphasic first-order kernel at a single pixel at the center of the RF from cat LGN X cell response to a spatiotemporal m-sequence. A segment of the m-sequence stimulus (at 128 Hz) at the center pixel is shown in Figure 2a, and the corresponding firing rate (in 7.7 ms bins) is shown in Figure 2b. The kernel estimate for the center pixel computed over the entire trial is shown in Figure 2c. The band represents two standard deviations around the estimator (for a discussion of the kernel estimator statistics, see section A.4). The uncertainty in the estimation (due to noise and unmodeled dynamics) is an extremely valuable measure, especially when comparing the encoding properties in different physiological states. Using the complete spatiotemporal kernel, the firing rate of the cell is predicted in Figure 2d; the actual firing rate is shown in the dark-shaded (positive) region, and the predicted firing rate is shown in the gray-shaded region, reflected about the horizontal axis for comparison. The cascade of the linear system with the

static nonlinearity is obviously a good predictor of the neuronal response for this relatively linear cell.

**2.2 Adaptive Estimation.** As shown in section A.3, a recursive estimate of the kernel at time  $t$  is posed as the least-squares estimate based on information up to time  $t$ :

$$\hat{\theta}_t = \arg \min_{\theta} \sum_{k=L}^t (r[k] - [\theta^T \varphi[k] + \mu_x])^2.$$

Kernel estimation formulated in this manner has various computational advantages (Stanley, 2000a; Ringach, Hawken, & Shapley, 2002). If this estimation is carried out over the entire trial, the resulting kernel will be equivalent to that of the nonrecursive estimate from equation 2.1, but the computation will be significantly more efficient. Perhaps the most beneficial aspect of the recursive estimation method, however, is the natural extension to adaptive estimation for neurons whose encoding dynamics vary over time, which has not been explored until recently (Stanley, 2000b, 2001; Brown, Nguyen, Frank, Wilson, & Solo, 2001). Many RF properties in the sensory pathways vary with time, violating a critical assumption for the reverse-correlation technique. Suppose that the recursive filter estimation (see section A.3) is reformulated as (Ljung, 1987),

$$\hat{\theta}_t = \arg \min_{\theta} \sum_{k=L}^t \lambda^{t-k} (r[k] - [\theta^T \varphi[k] + \mu_x])^2,$$

where  $\lambda \in [0, 1]$  is a weighting parameter. Let  $\Gamma_t$  and  $\Pi_t$  represent the input autocovariance and cross-covariance between input and output at time  $t$ , respectively, which can be expressed in an adaptive form:

$$\Gamma_t = \lambda \Gamma_{t-1} + \varphi[t] \varphi^T[t] \quad \Pi_t = \lambda \Pi_{t-1} + \varphi[t] r[t]. \quad (2.2)$$

As shown in section A.3, the estimate of the kernel at time  $t$  is finally written as the estimate at time  $t - 1$  plus a correction based on new information arriving at time  $t$ :

$$\begin{aligned} \hat{\theta}_t &= \hat{\theta}_{t-1} + \text{update} \\ &= \hat{\theta}_{t-1} + \frac{1}{t-L+1} \Gamma_t^{-1} \varphi[t] \left[ \frac{1}{C} r[t] - \varphi^T[t] \hat{\theta}_{t-1} \right]. \end{aligned} \quad (2.3)$$

This estimate therefore downweights past information in an exponential manner as controlled by the size of  $\lambda$ , which is often called the “forgetting” factor. As the encoding properties vary over time, the input-output covariance obviously changes properties, which is accounted for by a single weighting parameter, telescoping backward in time; the result is an

exponentially downweighting of past information. Let the time constant of the adaptive estimation  $\tau$  denote the time at which the exponential decay reaches 63% of its final value, which is  $\tau = \Delta t \times \ln(0.37) / \ln(\lambda)$ . The statistical properties of the adaptive estimator are discussed in section A.4.

### 3 Time-Varying Dynamics in Sensory Systems

---

Although there are computational advantages to the estimation posed above, the primary advantage in characterizing neuronal encoding is the ability to track time-varying characteristics of the cell response properties over a single trial. The importance of the paradigm is well illustrated through the following examples involving both simulation and experimentally obtained data from the mammalian visual pathway.

**3.1 Simulations of Time-Varying Encoding Dynamics in the LGN and Cortex.** Neurons in the LGN of cats have been shown to exhibit temporal tuning properties that change over time (Sherman & Koch, 1986; Mukherjee & Kaplan, 1995). Over a relatively short time frame of seconds to minutes, the dynamics can change smoothly from a tonically firing state to a bursting state, which is thought to be modulated by neuromodulatory influences from the visual cortex and brain stem. Mukherjee and Kaplan analyzed the transfer function characteristics between the visual input and LGN cell activity and found that the tonic mode is associated with low-pass temporal filter characteristics, whereas the burst mode has more of a bandpass nature, as shown in Figure 3a. Such nonstationary behavior over experimental trials precludes the accurate characterization of encoding properties that change over the experimental trial. The adaptive approach presented here, however, is specifically designed to capture such phenomena. Consider the following example for which both noise-free and more physiologically realistic stochastic simulations are conducted. Let the firing rate of the neuron simply be a linear function of a spatially uniform but temporally varying stimulus,  $r[t] = [\theta_t^T \varphi[t]]$ . Figure 3a shows the temporal frequency characteristics of the neurons in the tonic (low-pass) and bursting (bandpass) states. The system can be transformed to the parameterized form of the first-order Wiener kernel,  $\theta$ , where  $\theta_{tonic}$  and  $\theta_{burst}$  represent the dynamics of the two modes in the time domain, as shown in Figure 3b. The tonic mode is associated with the slower rise to the peak, whereas the burst mode has a quicker transient and a long inhibitory rebound. In order to simulate the transition from tonic firing to bursting and back to tonic, the dynamics of the neuron were generated from a time-varying weighted average of the two models:

$$\theta_t = \alpha_{tonic}(t)\theta_{tonic} + \alpha_{burst}(t)\theta_{burst}.$$

The weighting functions  $\alpha_{tonic}(t)$  and  $\alpha_{burst}(t)$  are shown in Figure 3c. The adaptive approach was then applied to estimate the time-varying kernel,

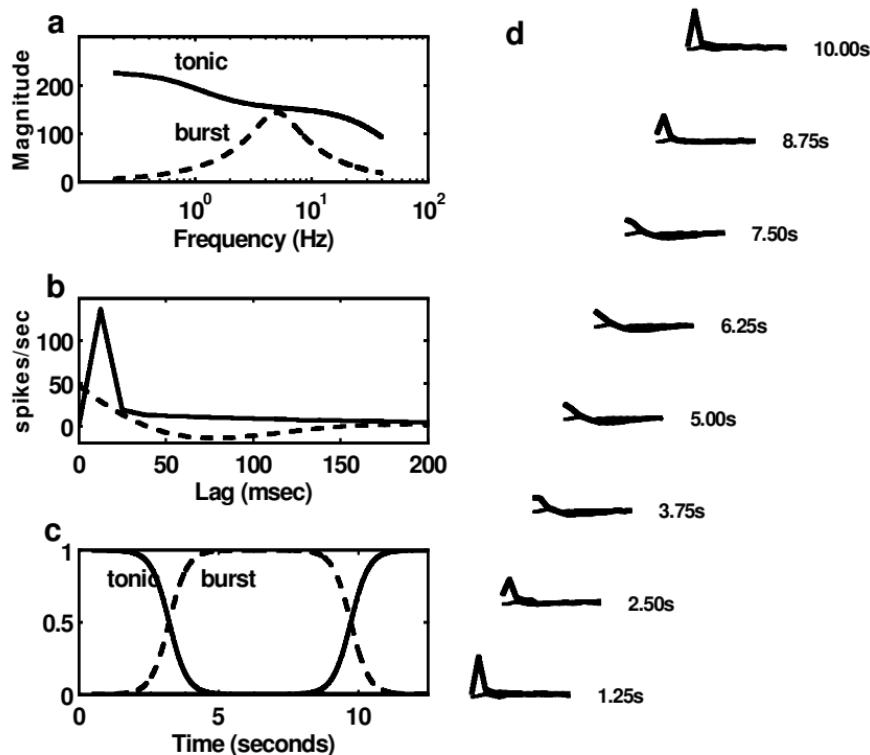


Figure 3: Adaptive tracking of temporal tuning modulation in LGN. (a) Temporal frequency response of tonic and burst firing modes and (b) associated kernels, (c) temporal weighting functions for tonic (solid) and burst (dashed) modes, and (d) kernels estimated during the course of the trial (thick line). The nonadaptive estimate is superimposed for reference (thin line). Kernel length was 175 msec and  $\tau$  was 800 msec.

and the adaptive time constant  $\tau$  was varied to produce the minimal prediction error. Figure 3d shows the first-order kernel estimate as a function of time, transitioning from the tonic to burst modes, and then back to the tonic mode over the time course of 10 seconds. For this case, the estimation time constant was approximately 800 msec. The estimate of the nonadaptive kernel is superimposed (thin line) for comparison. The transition from tonic to burst and then back to tonic modes is well captured in the kernel estimate for this noise-free simulation. The same analysis was then performed for the case when the rectified output of the linear kernel was used as the time-varying rate of an inhomogeneous Poisson process. A thinning algorithm was used to efficiently produce spike times from this construct (Dayan & Abbott, 2001); a segment of the stimulus and the spike train are

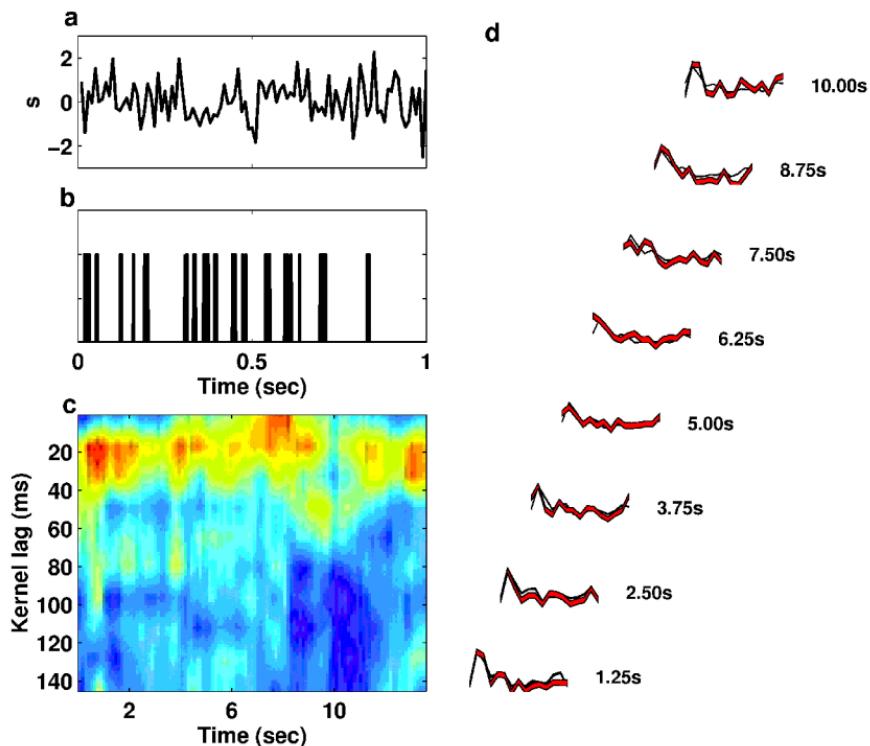


Figure 4: Adaptive tracking of temporal tuning modulation in LGN with noise. Temporal white noise (a) was filtered with the time-varying linear kernel, rectified, and then used as the rate of an inhomogeneous Poisson process to generate spikes (b). The adaptive algorithm tracked the time-varying kernel, as shown in the image (c), and the kernel evolution (black) in (d) where the nonadaptive kernel (gray) is again superimposed for reference. Bands represent two standard deviations around the estimate (see equation A.3). Kernel length was 175 msec, and  $\tau$  was approximately 800 msec.

shown in Figures 4a and 4b. The resulting spikes were binned to represent firing rate again, and the estimation procedure was implemented. An image of the kernel evolution over the stimulus trial is shown in Figure 4c, where the prominent peak of the tonic mode is present at the beginning and end of the trial, and the quick transients and inhibitory lobe of the burst mode are present in the middle of the trial. Figure 4d provides another view of the progression, again with adaptive estimate (red) and nonadaptive estimate (black) superimposed for comparison. Bands represent two standard deviations around the estimate (details of statistical properties are in section A.4). Although the estimate suffers as compared to the noise-free

simulations of Figure 3, the fundamental properties of the two modes are still extracted from the data set, which is clear from both illustrations. The nonadaptive estimate essentially averages out the interesting time-varying dynamics, while the adaptive estimate deviates from this over the 3.75 to 7.5 second range. The quick transient of burst mode is especially prominent in the adaptive estimate at 6.25 seconds. The confidence bands suggest that the adaptive estimates are significantly different from the nonadaptive estimates and also that the apparent evolution of the kernel properties over the trial is statistically significant.

The spatial RF properties of neurons in the visual cortex have been shown to exhibit plasticity not only over long timescales of days or months, but also over short timescales of seconds to minutes. It has been proposed that

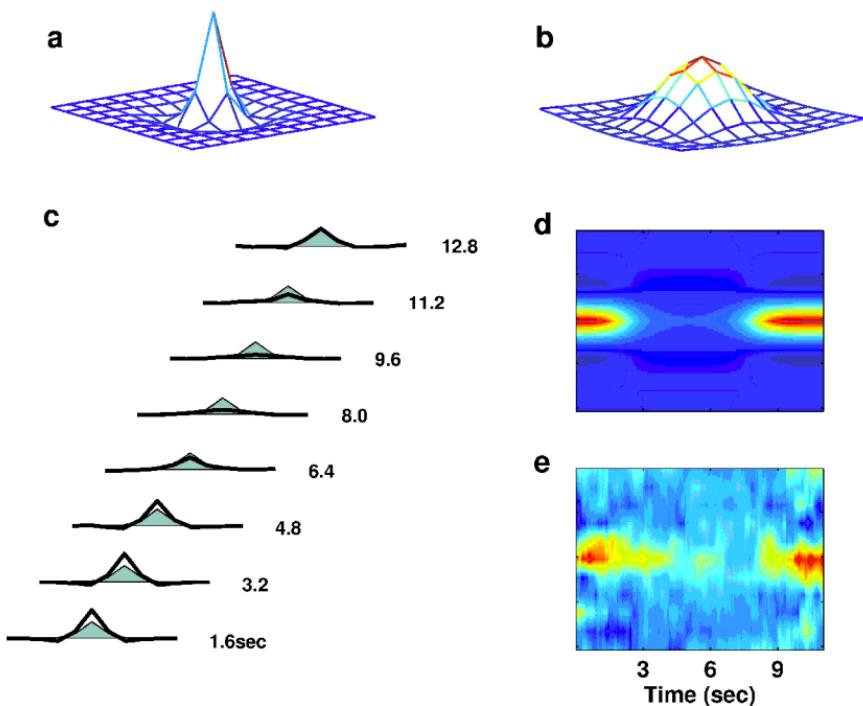


Figure 5: Dynamic cortical RF properties. (a) Original and (b) expanded RFs were used to simulate neuronal response, which was tracked adaptively (c). Thick lines represent the adaptively estimated spatial RF (over one dimension), and shaded regions represent the nonadaptive estimates. (d) An image of the RF evolution. In an additional simulation, the Wiener system output was used as the time-varying rate of an inhomogeneous Poisson process, producing a noisy response. The resulting RF properties (again for 1D) are shown for the adaptively estimated RF (e).  $\tau$  was approximately 800 msec.

this short-term plasticity plays a role in the continuing process of normalization and calibration of the visual system (Gilbert et al., 1996). Many cortical neurons exhibit stimulus-dependent RF expansions over relatively short time periods. Consider the following example, as illustrated in Figure 5. Concentrated and expanded RF were generated using difference of gaussian functions with different spatial extents, as shown in Figures 5a and 5b, with essentially instantaneous temporal dynamics. The time-varying system was then generated using the temporal weighting functions of the previous example, varying from the original to expanded and back to original RF structure. The response to spatiotemporal white noise was simulated, as previously discussed, resulting in a mean firing rate at 35 Hz. Figure 5c shows a spatial slice of the adaptively estimated 2D receptive field as it evolves over the trial (thick line) with the static RF estimate (shaded) superimposed for reference. The static RF estimate is essentially the averaged RF over the trial, whereas the actual RF was initially more peaked, followed by a broadening of the tuning, and finally a return to the peaked tuning at the end of the trial. The adaptive approach in this case is able to track these changes as they evolve over the stimulus trial, as shown in Figure 5d. In a more realistic simulation, the output of the linear filter-static nonlinearity cascade was used as the rate of an inhomogeneous Poisson process, as was done in the previous example. An image of the adaptively estimated RF (again for a 1D slice of the 2D field) over the stimulus trial is shown in Figure 5e. As with the previous example, the inherent noise of the Poisson simulation degrades the estimate, but the essential time-varying characteristics are still well captured.

**3.2 Experimental Results from LGN Response to the Spatiotemporal M-Sequence.** The above simulations provide insight into the development of the methods of adaptive estimation, but the primary target of the approach is experimental. X cells in the cat LGN were stimulated with the

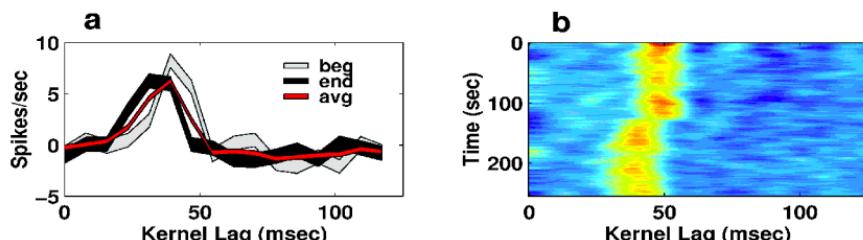


Figure 6: Evolution of adaptation in the LGN. (a) The first-order kernel at the beginning (beg) and end (end) of the 4 minute trial. For comparison, the kernel derived from the entire trial is superimposed (avg). Bands represent two standard deviations around the estimators. (b) An image of the kernel evolution over the 4 minute trial.  $\tau$  was 7.8 sec.

spatiotemporal m-sequence at 128 Hz and 100% contrast over several minutes (for details of experimental protocol, see Stanley, Li, & Dan, 1999). Figure 6 illustrates typical results obtained using the adaptive estimation approach for a single pixel at the center of the RF. Figure 6a shows the first-order kernel estimates at the beginning (beg) and end (end) of the trial, with a marked change in temporal dynamics over the 4 minute period. This example was typical among the population from which this cell was drawn, undergoing an attenuation over the stimulus trial, consistent with slow adaptation effects related to a step up in contrast at stimulus onset. Also invariant among cells is the temporal compression that results in a decreased latency. The kernel derived from the entire trial (avg) is superimposed for reference, underscoring the importance of the adaptive approach presented here; when the kernel is computed from the entire trial, it produces essentially the average dynamics over the trial. Bands represent two standard deviations around the estimators. The image shown in Figure 6b illustrates the evolution of the kernel dynamics over the entire trial, with a decrease in latency and in kernel magnitude over the 4 minute period. The adaptive time constant  $\tau$  was 7.8 seconds for this example. The sudden shift in latency is an artifact of the discretization of the kernel at 128 Hz, since the average shift in latency was on the order of magnitude of the sampling interval. The point to emphasize here is that these changes in encoding dynamics would not be discovered through traditional reverse-correlation techniques that assume a fixed relationship between stimulus and response over the trial.

#### 4 Discussion

---

Reverse-correlation techniques have been used extensively in a wide range of sensory systems. Characterizing the RF properties of sensory neurons forms the basis for our current understanding of their functional significance. The assumption that the spatiotemporal RF properties are static over even short time intervals, however, is problematic. Systems-level adaptation, functional modulation, and plasticity occur on a continuum of timescales that exist in all but the most artificial of experimental conditions. Examples were presented here in which spatiotemporal RF properties of visual neurons were adaptively tracked over short time intervals through both simulation and experimental trials. Realistic, noisy simulations were conducted in the development of the adaptive approach, shown through both modulation of temporal coding properties in the LGN and modulation of spatial properties in the cortex. Furthermore, experimental data were used to demonstrate the validity of the approach under physiological conditions, and it was shown that the evolution of encoding properties could be smoothly tracked over short trials. The kernels showed a marked attenuation and temporal compression over a timescale that is consistent with slow forms of contrast and light adaptation. Obviously this could be controlled

for experimentally by allowing the cell to adapt first to the high-contrast stimulus, but one could argue that this induces an unnatural steady state that is not relevant ethologically. Statistical properties of both the adaptive and nonadaptive estimators were derived, providing a more rigorous means with which to compare encoding dynamics in different physiological conditions. It should also be noted that the approach presented here tracks the evolution of spatiotemporal encoding properties over a single trial, an especially critical point in tracking dynamics that are modulated by nonstimulus-related activity, and thus may not be possible to repeat. Furthermore, for stimulus-related modulation of encoding dynamics, the applicability over single trials is important in natural viewing conditions, in which case the eyes saccade across the visual scene, never to invoke the same trajectory, and therefore stimulus pattern, twice.

In this work, the neuronal response is characterized as the single trial firing rate of the neuron in question. This simplification obviously ignores the precise temporal structure of the underlying point events that may be present in sparse coding strategies. However, for some subset of cells in early sensory systems, much of the information is thought to be coded in the firing rate. As is evident in Figure 2, X cells in the LGN are typically relatively linear in their response properties and are often well characterized with this strategy. The simple model prediction of firing rate is surprisingly good, although it may not be capturing subtle features that could be tracked through more elaborate strategies. One would expect that the specific techniques presented here would be less ideal for the sparse coding observed in many cortical cells. The relatively high firing rate of thalamic neurons enables the use of an adaptive time constant that is small enough to allow tracking of relatively fast changes in the RF properties. The lower firing rates associated with sparse coding strategies in the cortex would require much longer adaptive time constants, thereby limiting the effective bandwidth of adaptive processes that could be tracked. However, the general approach of adaptive estimation could certainly be applied; adaptive point process estimation techniques have been recently demonstrated in the hippocampus (Brown et al., 2001) and represent a promising direction of exploration in this context.

The adaptive approach presented here is based on an exponential down-weighting of past stimulus-response information. This functional form of the downweighting was chosen for computational reasons, but could easily be extended to more complex relationships. However, when this strategy was implemented with other functional forms of the temporal dependence, the results were qualitatively similar. The relevant parameter in each case was the relative temporal window over which the estimation was performed (the time constant of the adaptive estimate,  $\tau$ ). The shorter window obviously enables one to track dynamics that change more rapidly over time, but at the expense of the estimator variance; conversely, larger time constants reduce the variance but make it impossible to detect the nonstationarities.

The time constants used in this study were thus chosen so as to use as little past information as possible while still yielding meaningful structure in the kernel estimates.

One of the limitations of the proposed methodology lies in the underlying assumptions imposed on the model structure. The input to the static nonlinearity was assumed to be gaussian in nature. However, even with binary inputs, such as that of the m-sequence, the output of the linear block is nearly gaussian for even relatively short filter lengths. The nonlinearity described in this work is a simple half-wave rectification. This obviously does not capture saturation at high firing rates, but instead assumes that stimuli are sufficiently small as to avoid driving the cells to saturation. The approach presented here could certainly be integrated with recent techniques in the identification of arbitrary nonlinearities in the cascade (Chichilnisky, 2001; Nykamp & Ringach, 2001), in order to capture the phenomena over a larger operating range. Finally, it was also assumed that the mean of the input to the static nonlinearity was much smaller than the standard deviation (see section A.4), which significantly reduced the complexity of several of the estimates. However, as demonstrated in section A.2, this assumption was reasonable for the experimental data presented.

In summary, time-varying encoding properties may be a ubiquitous characteristic of all sensory neurons. Adaptation has been studied for some time and is known to dramatically affect the temporal and spatial dynamics, and anecdotal evidence of modulatory effects from other brain regions on encoding properties has been reported for a number of stages in the visual pathway. However, it is not yet known what these phenomena imply for the coding strategies of the sensory pathway as a whole. The approach presented here is a first step toward capturing the evolution of spatiotemporal RF properties over a range of timescales, with the goal of eventually understanding dynamic coding strategies employed by the visual pathway in natural settings, where there is no real concept of steady state.

## Appendix

---

**A.1 Time-Domain Estimation.** Consider a purely linear system, with zero-mean input  $s[k]$ , output  $x[k]$ , and first-order kernel  $g[k]$  that can be expressed through the vector  $\theta$ . Define a data vector  $\varphi$ ,

$$\varphi[k] \triangleq [s[k] \ s[k-1] \cdots s[k-L+1]]^T \in \mathbb{R}^{L \times 1}.$$

The output can be written  $x[k] = \theta^T \varphi[k]$ . Let the error in the prediction be defined as  $e[k] = x[k] - \hat{x}[k]$ . The mean square error (MSE) is then defined as  $MSE = E\{(e[k])^2\}$ , where  $E\{\cdot\}$  denotes statistical expectation. The first-order kernel can then be estimated by minimizing the MSE of prediction:

$$\hat{\theta} = \arg \min_{\theta \in \mathbb{R}^L} \sum_k \left\{ (r[k] - \theta^T \varphi[k])^2 \right\}.$$

Taking the partial of this objective function with respect to  $\theta[q]$  and setting equal to zero yields  $\phi_{sx}[q] = \sum_{m=0}^{L-1} \theta[m] \phi_{ss}[q-m]$ , where  $\phi_{sx}$  is the cross-covariance between the input and output, and  $\phi_{ss}$  is the autocovariance of the input. This can be expressed as:

$$\begin{bmatrix} \phi_{sx}[0] \\ \phi_{sx}[1] \\ \vdots \\ \phi_{sx}[L-1] \end{bmatrix} = \begin{bmatrix} \phi_{ss}[0] & \phi_{ss}[1] & \dots & \phi_{ss}[L-1] \\ \phi_{ss}[1] & \phi_{ss}[0] & \dots & \phi_{ss}[L-2] \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{ss}[L-1] & \phi_{ss}[L-2] & \dots & \phi_{ss}[0] \end{bmatrix} \begin{bmatrix} \theta[0] \\ \theta[1] \\ \vdots \\ \theta[L-1] \end{bmatrix}.$$

Using compact notation reduces the relationship to  $\phi_{sx} = \Phi_{ss}\theta$  where  $\Phi_{ss}$  is the  $L \times L$  Toeplitz matrix of the input autocovariance. The first-order kernel can then be estimated as  $\hat{\theta} = \Phi_{ss}^{-1}\phi_{sx}$ . This estimator is the optimal estimator in the least-squares sense (Ljung, 1987).

**A.2 Effect of Static Nonlinearity on Covariance Structure.** The cascade of a linear system with a static nonlinearity is often referred to as a Wiener system, as shown in Figure 1. It has been shown that the effect of odd static nonlinearities is a simple scaling of the covariance structure (Bussgang, 1952; Papoulis, 1984; Hunter & Korenberg, 1986). Bussgang's theorem states that if the input to a memoryless system  $r = f(x)$  is a normal process  $x(t)$ , the cross-covariance of  $x(t)$  with the output  $r(t)$  is proportional to the autocovariance of the input (Papoulis, 1984):

$$\phi_{xr}(\tau) = C\phi_{xx}(\tau) \quad \text{where} \quad C = E\{f'[x(t)]\}.$$

For a half-wave rectification, we have:

$$f(x) = \begin{cases} x & \text{for } x \geq 0 \\ 0 & \text{else} \end{cases} \quad f'(x) = \begin{cases} 1 & \text{for } x \geq 0 \\ 0 & \text{else} \end{cases}.$$

The scaling constant  $C$  then becomes

$$C = E\{f'(x)\} = \int_{-\infty}^{\infty} f'(x)p_x(x)dx,$$

where  $p_x(x)$  is assumed a gaussian probability density function  $\mathcal{N}(\mu_x, \sigma_x^2)$ . We therefore have

$$C = \int_0^{\infty} p_x(x)dx = \int_{\frac{-\mu_x}{\sigma_x}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = 1 - \Psi\left(\frac{-\mu_x}{\sigma_x}\right),$$

where  $\Psi(\cdot)$  is the standard normal cumulative. The theoretical and observed scalings for varying  $\mu_x/\sigma_x$  are shown in Figure 7. Gaussian white noise processes with mean  $\mu_x$  and variance  $\sigma_x^2$  were rectified; the ratio of cross-covariance to autocovariance,  $C$ , is plotted as a function of  $\mu_x/\sigma_x$  in Figure 7a.

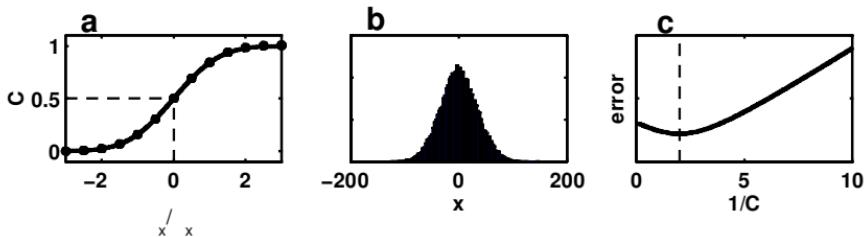


Figure 7: Effect of rectification on correlation structure. (a) Simulated and theoretical relationships between  $\mu_x/\sigma_x$  and scaling  $C$ . Gaussian white noise signals with mean  $\mu_x$  and variance  $\sigma_x^2$  were rectified, and the ratio of cross-covariance to autocovariance computed ( $\bullet$ ). (b) The output of the linear filter tends to a gaussian distribution. The nonscaled estimated kernel underpredicts the experimentally observed firing rate. (c) The  $\frac{1}{C}$  scaling on the kernel that minimizes the mean squared error is approximately 2, implying a  $\mu_x \ll \sigma_x$ , as indicated in a.

Importantly, data from the cat LGN reveal that the scaling of the kernel,  $1/C$ , which minimizes the error in predicted firing rate, is approximately 2, as shown in Figure 7c. The combination of these two results suggests that the mean of the filter output is small relative to the standard deviation. For the Wiener system shown in Figure 1, it is straightforward to extend the discussion above to the estimate for the linear block. For an uncorrelated input, the output of the linear kernel tends to a gaussian distribution, as shown in Figure 7b. The resulting relationship is  $\phi_{sr} = C\phi_{sx}$ , giving, for half-wave rectification, the estimate

$$\hat{\theta} = \frac{1}{C}\Phi_{ss}^{-1}\phi_{sr}.$$

Importantly, data from the cat LGN reveal that the scaling of the kernel,  $1/C$ , that minimizes the error in predicted firing rate is approximately 2, as shown in Figure 7c. The combination of these results suggests that  $\mu_x \ll \sigma_x$  within this structural framework.

**A.3 Recursive Estimation.** The recursive kernel estimation problem can be formulated in the following manner. After presenting a stimulus and observing the response of the neuron up to time  $t$ , the estimation can be expressed as

$$\hat{\theta}_t = \arg \min_{\theta} \sum_{k=L}^t (r[k] - [\theta^T \varphi[k] + \mu_x])^2,$$

where the subscript  $t$  denotes the time-varying nature of the parameter vector  $\theta$ . Using the least-squares technique described previously, the filter

can be estimated based on information up to time  $t$ ,

$$\hat{\theta}_t = \frac{1}{C} \Gamma_t^{-1} \Pi_t,$$

where  $\Gamma_t$  and  $\Pi_t$  are the stimulus autocovariance Toeplitz matrix and cross-covariance between stimulus and response, respectively, based on data up to time  $t$ :

$$\begin{aligned}\Gamma_t &= \frac{1}{t-L+1} \sum_{k=L}^t \varphi[k] \varphi^T[k] = \frac{t-L}{t-L+1} \Gamma_{t-1} + \frac{1}{t-L+1} \varphi[t] \varphi^T[t] \\ \Pi_t &= \frac{1}{t-L+1} \sum_{k=L}^t \varphi[k] r[k] = \frac{t-L}{t-L+1} \Pi_{t-1} + \frac{1}{t-L+1} \varphi[t] r[t].\end{aligned}$$

This gives the final expression for the estimate:

$$\hat{\theta}_t = \hat{\theta}_{t-1} + \frac{1}{t-L+1} \Gamma_t^{-1} \varphi[t] \left[ \frac{1}{C} r[t] - \varphi^T[t] \hat{\theta}_{t-1} \right].$$

The estimate based on data up to time  $t-1$  is therefore used in the subsequent estimate based on information up to time  $t$ , which results in a recursion for the filter estimation (Goodwin & Sin, 1984; Ljung & Söderström, 1983).

**A.4 Statistical Properties of Estimates.** As with the traditional least-squares problem, the statistics of the estimator for the Wiener system can be characterized in a relatively straightforward manner. It is first necessary to establish the statistical properties of the linear problem, given here for reference. Define the following data matrices:

$$D \triangleq [\varphi[L] \quad \varphi[L+1] \quad \cdots \quad \varphi[N]]^T \in \mathbb{R}^{(N-L+1) \times L}$$

$$\psi \triangleq [x[L] \quad x[L+1] \quad \cdots \quad x[N]]^T \in \mathbb{R}^{(N-L+1) \times 1}.$$

The estimate of the parameter vector from section A.1 can then be written in a more standard least-squares form,

$$\hat{\theta} = \Gamma^{-1} \Pi = (D^T D)^{-1} D^T \psi,$$

where again  $\Gamma$  represents the stimulus autocovariance Toeplitz matrix, and  $\Pi$  represents the cross-covariance matrix between stimulus and response. This estimator is known to be unbiased, so that  $E[\hat{\theta}] = \theta$ . The covariance can be written (Westwick, Suki, & Lutchen, 1998)

$$\Lambda_{\hat{\theta}} = E\{(\hat{\theta} - \theta)(\hat{\theta} - \theta)^T\} = (D^T D)^{-1} D^T E\{ee^T\} D (D^T D)^{-1},$$

where  $e = [e[L] \ e[L+1] \ \cdots \ e[N]]^T \in \mathbb{R}^{N-L+1 \times 1}$  is the noise vector, where  $e[k] \triangleq x[k] - \hat{x}[k]$ . If the noise is a white process, with standard deviation  $\sigma_e \gg \mu_e$ , then  $E\{ee^T\} \approx \sigma_e^2 I$ , resulting in

$$\Lambda_{\hat{\theta}} \approx \sigma_e^2 (D^T D)^{-1} = \frac{\sigma_e^2}{N-L+1} \Gamma^{-1} \in \mathbb{R}^{N-l+1 \times N-L+1}.$$

The covariance structure of the estimator  $\hat{\theta}$  could be directly obtained from the noise term  $e[k] = x[k] - \hat{x}[k]$ . In this case, however, there is obviously no access to the intermediate signal  $x[k]$ . The unobservable noise term  $e$  must therefore be related to observable noise, which is the difference between the actual firing rate of the cell  $r[k]$  and that predicted by the linear-nonlinear cascade,  $n[k] = r[k] - \hat{r}[k]$ .

In order to understand how the two noise processes relate, it is helpful to consider first the effect of the static nonlinearity on the probability density function. Suppose  $x$  is a random variable such that  $x \sim \mathcal{N}(\mu_x, \sigma_x^2)$ . Let  $p_x(x)$  represent the density function associated with the random variable  $x$ . Let  $r$  be the output of the static nonlinearity  $f(\cdot)$  such that  $r = f(x)$ . For the half-wave rectification, the expectation becomes

$$\mu_r = E\{r\} = \int_{-\infty}^{\infty} rp_r(r) dr = \int_{-\infty}^{\infty} f(x)p_x(x) dx = \int_0^{\infty} x \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{\frac{-(x-\mu_x)^2}{2\sigma_x^2}} dx.$$

Letting  $z = \frac{x-\mu_x}{\sigma_x}$ , the expression becomes:

$$\begin{aligned} E\{r\} &= \frac{1}{\sqrt{2\pi\sigma_x^2}} \int_{-\frac{\mu_x}{\sigma_x}}^{\infty} (\sigma_x z + \mu_x) e^{\frac{-z^2}{2}} \sigma_x dz \\ &= \sigma_x \int_0^{\infty} \frac{1}{\sqrt{2\pi}} ze^{\frac{-z^2}{2}} dz + \sigma_x \int_{-\frac{\mu_x}{\sigma_x}}^0 \frac{1}{\sqrt{2\pi}} ze^{\frac{-z^2}{2}} dz + \mu_x \int_{-\frac{\mu_x}{\sigma_x}}^{\infty} e^{\frac{-z^2}{2}} dz. \end{aligned}$$

If  $\mu_x \ll \sigma_x$ , the expectation becomes:

$$E\{r\} = \sigma_x \int_0^{\infty} \frac{1}{\sqrt{2\pi}} ze^{\frac{-z^2}{2}} dz = \frac{\sigma_x}{\sqrt{2\pi}}.$$

From the sample mean of the output signal  $r$ , the variance of the underlying zero mean noise process can be estimated. For the variance,  $\sigma_r^2 = E\{r^2\} - \mu_r^2$ , where

$$\begin{aligned} E\{r^2\} &= \frac{\sigma_x^3}{\sqrt{\sigma_x^2}} \int_{-\frac{\mu_x}{\sigma_x}}^{\infty} z^2 \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}} dz \\ &\quad + 2 \frac{\mu_x \sigma_x^2}{\sqrt{\sigma_x^2}} \int_{-\frac{\mu_x}{\sigma_x}}^{\infty} z \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}} dz + \frac{\mu_x^2 \sigma_x^2}{\sqrt{\sigma_x^2}} \int_{-\frac{\mu_x}{\sigma_x}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}} dz. \end{aligned}$$

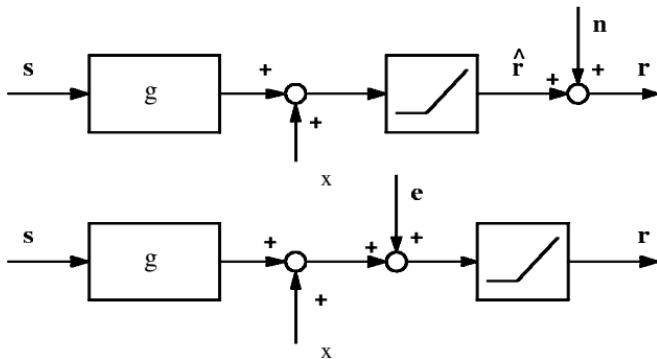


Figure 8: Equivalent noise sources. The top block diagram represents the encoding model as observed experimentally, with the source of noise at the point of measurement. The bottom block diagram represents the dynamics assumed in the adaptive estimation scheme. As detailed in the text, the statistical properties of the hidden noise source  $e$  can be estimated from the observed “noise”  $n$ .

If again  $\mu_x \ll \sigma_x$ , this results in

$$\sigma_r^2 = E\{r^2\} - E\{r\}^2 = \frac{\sigma_x^2}{2} - \frac{\sigma_x^2}{2\pi} = \sigma_x^2 \frac{(\pi - 1)}{2\pi}.$$

This result can now be extended to describe the equivalent noise sources shown in Figure 8. For the top block diagram,

$$\sigma_r^2 = \sigma_{\hat{r}}^2 + \sigma_n^2 = \frac{\pi - 1}{2\pi} \sigma_x^2 + \sigma_n^2 = \frac{\pi - 1}{2\pi} \|\theta\|^2 \sigma_s^2 + \sigma_n^2.$$

Similarly, for the bottom system,

$$\sigma_r^2 = \frac{\pi - 1}{2\pi} \sigma_x^2 = \frac{\pi - 1}{2\pi} (\sigma_{\hat{x}}^2 + \sigma_e^2) = \frac{\pi - 1}{2\pi} (\|\theta\|^2 \sigma_s^2 + \sigma_e^2),$$

where  $\|\cdot\|$  is the standard Euclidean norm. Putting the two equations together yields

$$\sigma_n^2 = \frac{\pi - 1}{2\pi} \sigma_e^2.$$

This gives the equivalent variance of the two noise sources. Given the observed neuronal response  $r$ , the statistical properties of the hidden noise process  $e$  can be estimated, which is necessary for determining the statistics of the least-squares estimate. The estimator remains unbiased, and the covariance of the estimator can be written as

$$\Lambda_{\hat{\theta}} = \sigma_e^2 (D^T D)^{-1} = \frac{\sigma_e^2}{N - L + 1} \Gamma^{-1} = \frac{1}{N - L + 1} (2\pi \mu_r^2 - \|\theta\|^2 \sigma_s^2) \Gamma^{-1}, \quad (\text{A.1})$$

where  $\mu_r$  is the mean observed firing rate and  $\sigma_s^2$  is the variance of the stimulus. The quantities  $\mu_r$ ,  $\theta$ , and  $\sigma_s^2$  can obviously be estimated from data. The uncertainty in the kernel estimate in Figure 2 was computed in this manner. This argument extends naturally to the recursive estimates, for which the covariance becomes

$$\Lambda_{\hat{\theta}_t} = \frac{1}{t-L+1} (2\pi\mu_r^2 - \|\theta\|^2 \sigma_s^2) \Gamma^{-1}. \quad (\text{A.2})$$

Within the adaptive framework, the estimate remains unbiased, as with the nonweighted estimation problem, but the covariance of the estimator has a slightly different form. The previous expression for covariance was normalized by the data length used in the estimate. With the adaptive algorithm presented here, the normalization depends on the structure of the exponential decay of contributions from the past:

$$\Lambda_{\hat{\theta}_t} = \frac{1}{w} (2\pi\mu_r^2 - \|\theta\|^2 \sigma_s^2) \Gamma_t^{-1} \quad w \triangleq \sum_{k=L}^t \lambda^{t-k+1}, \quad (\text{A.3})$$

where  $w$  reflects the exponential downweighting of past information. Note that if  $\lambda = 1$ , the estimator variance matches that of the nonadaptive estimator. The uncertainty in the kernel estimate in Figures 4 through 6 were computed in this manner.

## Acknowledgments

---

I thank Emery Brown, Roger Brockett, and Nick Lesica for helpful comments in the preparation of this article and Yang Dan for the use of the physiological data.

## References

---

- Albrecht, D. G., Farrar, S. B., & Hamilton, D. B. (1984). Spatial contrast adaptation characteristics of neurones recorded in the cat's visual cortex. *J. Physiol.*, *347*, 713–739.
- Brown, E. N., Nguyen, D. P., Frank, L. M., Wilson, M. A., & Solo, V. (2001). An analysis of neural receptive field plasticity by point process adaptive filtering. *PNAS*, *98*, 12261–12266.
- Bussgang, J. J. (1952). Crosscorrelation functions of amplitude-distorted gaussian signals. *MIT Res. Lab. Elec. Tech. Rep.*, *216*, 1–14.
- Chichilnisky, E. J. (2001). A simple white noise analysis of neuronal light responses. *Network*, *12*, 199–213.
- Dan, Y., Atick, J., & Reid, R. (1996). Efficient coding of natural scenes in the lateral geniculate nucleus: Experimental test of a computation theory. *J. Neurosci.*, *16*(10), 3351–3362.

- Dayan, P., & Abbott, L. F. (2001). *Theoretical neuroscience*. Cambridge, MA: MIT Press.
- DeBoer, E., & Kuyper, P. (1968). Triggered correlation. *IEEE Trans. Biomed. Eng.*, 15, 169–179.
- deCharms, R. C., Blake, D., & Merzenich, M. (1998). Optimizing sound features for cortical neurons. *Science*, 280, 1439–1443.
- Enroth-Cugell, C., & Shapley, R. (1973). Adaptation and dynamics of cat retinal ganglion cells. *J. Physiol.*, 233, 271–309.
- Gilbert, C. D., Das, A., Ito, M., Kapadia, M., & Westheimer, G. (1996). Spatial integration and cortical dynamics. *Proc. Natl. Acad. Sci. USA*, 93, 615–622.
- Goodwin, G. C., & Sin, K. S. (1984). *Adaptive filtering prediction and control*. Upper Saddle River, NJ: Prentice Hall.
- Greblicki, W. (1992). Nonparametric identification of Wiener systems. *IEEE Trans. on Inform. Thry.*, 38, 1487–1493.
- Greblicki, W. (1994). Nonparametric identification of Wiener systems by orthogonal series. *IEEE Trans. on Automatic Control*, 39, 2077–2086.
- Greblicki, W. (1997). Nonparametric approach to Wiener system identification. *IEEE Trans. on Circuits and Sys.*, 44, 538–545.
- Heeger, D. (1992). Normalization of cell responses in cat striate cortex. *Vis. Neurosci.*, 9, 181–197.
- Hunter, I. W., & Korenberg, M. J. (1986). The identification of nonlinear biological systems: Wiener and Hammerstein cascade models. *Biological Cybernetics*, 55, 135–144.
- Jones, J. P., & Palmer, L. A. (1987). The two-dimensional spatial structure of simple receptive fields in cat striate cortex. *J. Neurophysiol.*, 58, 1187–1211.
- Korenberg, M. J., & Hunter, I. W. (1999). Two methods for identifying Wiener cascades having noninvertible static nonlinearities. *Annals of Biomed. Eng.*, 27, 793–804.
- Ljung, L. (1987). *System identification: Theory for the user*. Upper Saddle River, NJ: Prentice Hall.
- Ljung, L., & Söderström, T. (1983). *Theory and practice of recursive identification*. Cambridge, MA: MIT Press.
- Mao, B. Q., MacLeish, P. R., & Victor, J. D. (1998). The intrinsic dynamics of retinal bipolar cells isolated from tiger salamander. *Vis. Neurosci.*, 15, 425–438.
- Movshon, J. A., Thompson, I. D., & Tolhurst, D. J. (1978). Spatial summation in the receptive fields of simple cells in the cat's striate cortex. *J. Physiol.*, 283, 53–77.
- Mukherjee, P., & Kaplan, E. (1995). Dynamics of neurons in the cat lateral geniculata nucleus: In vivo electrophysiology and computational modeling. *J. Neurophysiol.*, 74(3), 1222–1243.
- Nykamp, D. Q., & Ringach, D. L. (2001). Full identification of a linear-nonlinear system via cross-correlation analysis. *Journal of Vision*, 2, 1–11.
- Ohzawa, I., Sclar, G., & Freeman, R. D. (1985). Contrast gain control in the cat's visual system. *Journal of Neurophysiology*, 54, 651–667.
- Papoulis, A. (1984). *Probability, random variables, and stochastic processes* (2nd ed.). New York: McGraw-Hill.

- Paulin, M. G. (1993). A method for constructing data-based models of spiking neurons using a dynamic linear-static nonlinear cascade. *Biol. Cybern.*, 69, 67–76.
- Pei, X., Vidyasagar, T. R., Volgushev, M., & Creutzfeld, O. D. (1994). Receptive field analysis and orientation selectivity of postsynaptic potentials of simple cells in cat visual cortex. *J. Neurosci.*, 14(11), 7130–7140.
- Przybyszewski, A. W., Gaska, J. P., Foote, W., & Pollen, D. A. (2000). Striate cortex increases contrast gain of macaque LGN neurons. *Visual Neuroscience*, 17, 485–494.
- Reid, R. C., Soodak, R. E., & Shapley, R. M. (1991). Directional selectivity and spatiotemporal structure of receptive fields of simple cells in cat striate cortex. *J. Neurophysiol.*, 66, 505–529.
- Reid, R. C., Victor, J. D., & Shapley, R. M. (1997). The use of m-sequences in the analysis of visual neurons: Linear receptive field properties. *Vis. Neurosci.*, 14, 1015–1027.
- Ringach, D. L., Hawken, M. J., & Shapley, R. (2002). Receptive field structure in neurons in monkey primary visual cortex revealed by stimulation with natural image sequences. *Journal of Vision*, 2, 12–24.
- Ringach, D., Sapiro, G., & Shapley, R. (1997). A subspace reverse-correlation technique for the study of visual neurons. *Vision Res.*, 37, 2455–2464.
- Shapley, R., & Victor, J. D. (1978). The effect of contrast on the transfer properties of cat retinal ganglion cells. *J. Physiol.*, 285, 275–298.
- Sherman, S., & Koch, C. (1986). The control of retinogeniculate transmission in the mammalian lateral geniculate nucleus. *Exp. Brain. Res.*, 63, 1–20.
- Smirnakis, S. M., Berry, M. J., Warland, D. K., Bialek, W., & Meister, M. (1997). Adaptation of retinal processing to image contrast and spatial scale. *Nature*, 386, 69–73.
- Stanley, G. B. (2000a). *Algorithms for real-time prediction in neural systems*. Paper presented at the Annual Meeting of the Engineering in Medicine and Biology Society, Chicago.
- Stanley, G. B. (2000b). *Time-varying properties of neurons in visual cortex: Estimation and prediction*. Paper presented at the 30th Annual Meeting of the Society for Neuroscience, New Orleans, LA.
- Stanley, G. B. (2001). Recursive stimulus reconstruction algorithms for real-time implementation in neural ensembles. *Neurocomputing*, 38–40, 1703–1708.
- Stanley, G., Li, F., & Dan, Y. (1999). Reconstruction of natural scenes from ensemble responses in the lateral geniculate nucleus. *Journal of Neuroscience*, 19(18), 8036–8042.
- Tolhurst, D. J., & Dean, A. F. (1990). The effects of contrast on the linearity of the spatial summation of simple cells in the cat's striate cortex. *Experimental Brain Research*, 79, 582–588.
- Westwick, D. T., Suki, B., & Lutchen, K. (1998). Sensitivity analysis of kernel estimates: Implications in nonlinear physiological system identification. *Annals of Biomend Eng.*, 26, 488–501.