

Submitted to: The Annual Computational Neuroscience Meeting (CNS*03)
Alicante, Spain, 5-9 July 2003.
Revision: 23 February 2003.

Maximum Likelihood Estimation of a Stochastic Integrate-and-Fire Neural Model*

Jonathan W. Pillow Liam Paninski Eero P. Simoncelli
pillow@cns.nyu.edu liam@cns.nyu.edu eero.simoncelli@nyu.edu

Howard Hughes Medical Institute, and
Center for Neural Science
New York University

Abstract

Recent work has examined the estimation of models of stimulus-driven neural activity in which some linear filtering process is followed by a nonlinear, probabilistic spiking mechanism. We analyze the estimation of one such model for which this nonlinear step is implemented by a noisy, leaky, integrate-and-fire mechanism. Specifically, we formulate the problem in terms of maximum likelihood estimation, and show that the computational problem of optimizing this cost function is tractable. Our main contribution is an algorithm and a proof that this algorithm is guaranteed to find the global optimum. We demonstrate the effectiveness of our estimator with numerical simulations.

A central issue in computational neuroscience is the characterization of the functional relationship between sensory stimuli and neural spike trains. A common model for this relationship consists of linear filtering of the stimulus, followed by a nonlinear, probabilistic spike generation process. The linear filter is typically interpreted as the neuron’s “receptive field,” while the spiking mechanism accounts for simple nonlinearities like rectification and response saturation. Given a set of stimuli and (extracellularly) recorded spike times, the characterization problem consists of estimating both the linear filter and the parameters governing the spiking mechanism.

A specific and widely used model of this type is the Linear-Nonlinear-Poisson (LNP) cascade model, in which spikes are generated according to an inhomogeneous Poisson process, with rate determined by an instantaneous (“memoryless”) nonlinear function of the filtered input. This model has a number of desirable features, including conceptual simplicity and computational tractability. In particular, reverse correlation analysis provides a simple unbiased estimator for the linear filter [e.g. 4], and the properties of estimators (for both the linear filter and static nonlinearity) have been thoroughly analyzed, even for the case of highly non-symmetric or “naturalistic” stimuli [9]. One important drawback of the LNP model, however, is that Poisson processes do not accurately capture the statistics of neural spike trains [2, 7, 12, 1]. In particular, the

* JWP and LP contributed equally to this work.
We thank E.J. Chichilnisky for helpful discussions.

probability of observing a spike is not a functional of the stimulus only; it is also affected by the recent history of spiking.

The leaky integrate-and-fire (LIF) model provides a biophysically more realistic spike mechanism, with a form of spike-history dependence. This model is simple, well-understood, and has dynamics that are entirely linear except for a nonlinear “reset” of the membrane potential following a spike. Although this model’s overriding linearity is often emphasized (owing to its approximately linear input current - frequency curve and lack of active conductances), the nonlinear reset has significant functional importance for the model’s response properties. In previous work, we have shown that standard reverse correlation analysis fails when applied to a neuron with deterministic (noise-free) LIF spike generation; we developed a new estimator for this model, and demonstrated that a change in leakiness of such a mechanism might underlie nonlinear effects of contrast adaptation in macaque retinal ganglion cells [11]. We and others have explored other “adaptive” properties of the LIF model [13, 10].

In this paper, we consider a model consisting of a linear filter followed by noisy LIF spike generation; this is essentially the standard LIF model driven by a noisy, filtered version of the stimulus. (We will refer to this as the L-NLIF model.) The probabilistic nature of this model provides several important advantages over the deterministic setup we have considered previously. First, an explicit noise model allows us to couch the problem in the terms of classical estimation theory. This, in turn, provides a natural “cost function” (likelihood) for model assessment and leads to more efficient estimation of the model parameters. Second, noise allows us to explicitly model neural firing statistics, and could provide a rigorous basis for a metric distance between spike trains, useful in other contexts [14]. Finally, noise influences the behavior of the model itself, giving rise to phenomena not observed in the purely deterministic model [8].

Our main contribution here is to show that the maximum likelihood estimator (MLE) for the L-NLIF model is computationally tractable. Specifically, we describe an algorithm for computing the likelihood function (involving multidimensional probability integrals), and prove that this likelihood function contains no local maxima, implying that the MLE can be computed efficiently using standard ascent techniques. The desirable statistical properties of this estimator (e.g. consistency, efficiency) are all inherited “for free” from classical estimation theory. Thus, we have a compact and powerful model for the neural code, and a well-motivated, efficient way to estimate the parameters of this model from extracellular data.

The Model

We consider a model for which the (dimensionless) subthreshold voltage variable V evolves according to

$$\frac{dV}{dt} = -gV(t) + \vec{k} \cdot \vec{x}(t) + \sigma n(t), \quad (1)$$

and resets to V_r whenever $V = 1$. Here, g denotes the leak conductance, $\vec{k} \cdot \vec{x}(t)$ the projection of the input signal $\vec{x}(t)$ onto the linear kernel \vec{k} , and $n(t)$ an unobserved

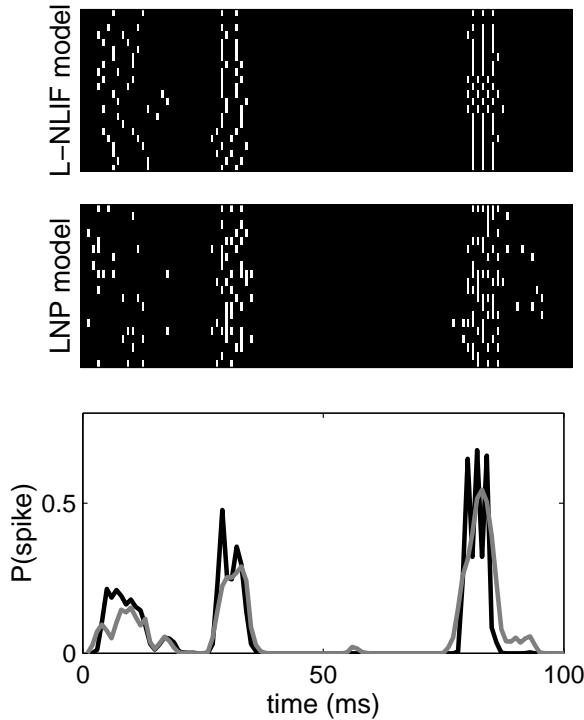


Figure 1: Simulated responses of L-NLIF and LNP models to 20 repetitions of a fixed 100-ms stimulus segment of temporal white noise. **Top:** Raster of responses of L-NLIF model, where $\sigma_{noise}/\sigma_{signal} = .5$, \vec{k} is a biphasic filter, and g gives a membrane time constant of 15 ms. The top row shows the fixed (deterministic) response of the model with σ_{noise} set to zero. **Middle:** Raster of responses of LNP model, with parameters fit with standard methods from a long run of the L-NLIF model responses to non-repeating stimuli. **Bottom:** (Black line) Post-stimulus time histogram (PSTH) of the simulated L-NLIF response. (Gray line) PSTH of the LNP model. Note that the L-NLIF preserves the fine temporal structure of the spike trains, relative to the LNP model.

(hidden) noise process with scale parameter σ . Without loss of generality, the “leak” and “threshold” potential are set at 0 and 1, respectively, so the cell spikes whenever $V = 1$, and decays back to 0 with time constant $1/g$ in the absence of input. Note that the nonlinear behavior of the model is completely determined by only a few parameters, namely $\{g, \sigma, V_r\}$. Figure 1 shows a simple comparison of the L-NLIF and LNP models. Note, in particular, the fine structure of spike timing in the responses of the L-NLIF model, which is qualitatively similar to *in vivo* experimental observations [2, 12, 7]).

Our problem now is to estimate the model parameters $\{\vec{k}, g, \sigma, V_r\}$ from a sufficiently rich, dynamic input sequence $\vec{x}(t)$ together with spike times t_i . A natural choice is the maximum likelihood estimator (MLE), which is easily proven to be consistent and statistically efficient here. To compute the MLE, we need to compute the likelihood and develop an algorithm for maximizing it.

The tractability of the likelihood function for this model arises directly from the linearity of the subthreshold dynamics of voltage $V(t)$ during an interspike interval. In the noiseless case [11], the voltage trace during an interspike interval $t \in [0, t_i]$ is given by the solution to equation (1) with $\sigma = 0$:

$$V(t) = V_r e^{-gt} + \int_0^t (\vec{k} \cdot \vec{x}(s)) e^{-gs} ds, \quad (2)$$

which is simply a linear convolution of the filtered stimulus with a negative exponential. It is easy to see that adding Gaussian noise to the voltage during each time step induces a Gaussian density over $V(t)$, since linear dynamics preserve Gaussianity [6]. This density is uniquely characterized by its first two moments; the mean is given by (2), and its covariance can easily be shown to be $E_g E_g^T$, where E_g is the convolution operator corre-

sponding to e^{-gt} . Note that this density is highly correlated for nearby points in time, since noise is integrated by the linear dynamics. Intuitively, smaller leak conductance g leads to stronger correlation in $V(t)$ at nearby time points. We denote this Gaussian density $G(\vec{x}_i, \vec{k}, \sigma, g, V_r)$, where index i indicates the i th spike and the corresponding stimulus chunk \vec{x}_i (i.e. the stimuli that influence $V(t)$ during the i th interspike interval).

Now, on any interspike interval $t \in [0, t_i]$, the only information we have is that $V(t)$ is less than threshold for all times before t_i , and exceeds threshold during the time bin containing t_i . This translates to a set of linear constraints on the discretized representation of $V(t)$, expressed in terms of the set C_i defined as:

$$C_i = \bigcap_{0 \leq t < t_i} \left\{ V(t) < 1 \right\} \cap \left\{ V(t_i) \geq 1 \right\}.$$

Therefore, the likelihood that the neuron first spikes at time t_i , given a spike at time 0, is the probability of the event $V(t) \in C_i$, which is given by the integral of the Gaussian density $G(\vec{x}_i, \vec{k}, \sigma, g, V_r)$ over the set C_i .

Spiking resets V to V_r , meaning that the noise contribution to V in different interspike intervals is independent. This, in turn, implies that the density over $V(t)$ for an entire experiment factorizes into a product of conditionally independent terms, where each of these terms is one of the Gaussian integrals derived above for a single interspike interval. The likelihood for the entire spike train is therefore the product of these terms over all observed spikes. Putting all the pieces together, then, define the likelihood as

$$L_{\{\vec{x}_i, t_i\}}(\vec{k}, g, \sigma, V_r) = \prod_i \int_{C_i} G(\vec{x}_i, \vec{k}, \sigma, g, V_r),$$

where the product, again, is over all observed spike times t_i and corresponding stimulus chunks \vec{x}_i .

Now that we have an expression for the likelihood, we need to be able to maximize it. Our main result now states, basically, that we can use simple ascent algorithms to compute the MLE without getting stuck in local maxima.

Theorem 1. *The likelihood $p_{\{\vec{x}_i, t_i\}}(\vec{k}, g, \sigma, V_r)$ has no local extrema in the parameters $(\vec{k}, g, \sigma, V_r)$, for any data $\{\vec{x}_i, t_i\}^1$.*

The proof is based on the logconcavity of $p_{\{\vec{x}_i, t_i\}}(\vec{k}, g, \sigma, V_r)$ under a certain parametrization of $(\vec{k}, \sigma, g, V_r)$. The classical approach for establishing the nonexistence of local maxima of a given function is concavity, which corresponds roughly to the function having everywhere non-positive second derivatives. However, the basic idea can be extended with the use of any invertible function: if f has no local extrema, neither will $g(f)$, for any strictly increasing real function g . The logarithm is a natural choice for g in any probabilistic context in which independence plays a role, since sums are easier to work

¹See <http://www.cns.nyu.edu/~liam/IFproof.html> for the full proof of this result

with than products. Moreover, concavity of a function f is strictly stronger than logconcavity, so logconcavity can be a powerful tool even in situations for which concavity is useless. (The Gaussian density is logconcave but not concave, for example). Our proof relies on a particular theorem [3] establishing the logconcavity of integrals of logconcave functions, and proceeds by making a correspondence between this type of integral and the integrals that appear in the definition of the L-NLIF likelihood above.

We should also note that the proof extends without difficulty to some other noise processes which generate logconcave densities (where white noise has the standard Gaussian density); for example, the proof is nearly identical if $n(t)$ is allowed to be colored Gaussian noise with nonzero drift. In this case the MLE also recovers the mean and covariance of the noise.

Computational methods and numerical results

Theorem 1 tells us that we can ascend the likelihood surface without fear of getting stuck in local maxima. Now how do we actually compute the likelihood? This is a nontrivial problem: we need to be able to quickly compute (or at least approximate, in a rational way) integrals of multivariate Gaussian densities G over simple but high-dimensional orthants C_i . While we lack the room for a systematic description of our solution to this problem, we can give the two main ingredients. The first is a specialized algorithm due to Genz [5], designed to compute exactly the kinds of integrals considered here, which works well when the boxes C_i are defined by fewer than ≈ 10 linear constraints. The number of actual constraints grows linearly in the length of the interspike interval t_i ; thus, to use this algorithm in typical data situations, we must adopt an approximative strategy of using only a subset of the constraints. We used a similar strategy in our previous work in the noiseless context [11]; the key point is that only a few constraints are needed to approximate the full integrals to a high degree of precision. The second ingredient, therefore, is a simple method for computing bounds on the error incurred by not using all of the available constraints, and a greedy algorithm for quickly finding the approximation which minimizes these bounds (and therefore efficiently picks out the most “important” constraints). An application to simulated data is shown in Fig. 2.

In conclusion, we have shown that a model which couples a linear filtering stage to a more realistic model of neuronal spiking, the L-NLIF model, can be tractably and efficiently estimated from extracellular physiological data using maximum likelihood. One limitation of this model is that it cannot be directly applied in cases where the NLIF spike generator is driven by a nonlinear transformation of the raw stimulus (e.g. complex cells in V1). However, the tractability and biological plausibility make the L-NLIF an attractive alternative to the LNP model, and to more complicated models with primarily linear drive. Our next step will be to apply the model to physiological data, to assess whether it accurately accounts for the stimulus preferences and spiking statistics of real neurons.

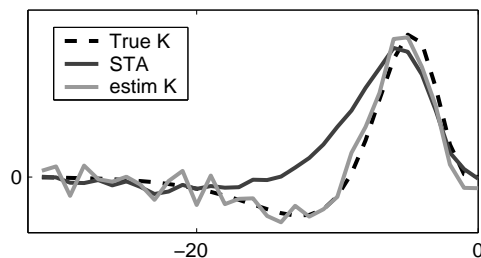


Figure 2: Demonstration of the estimator’s performance on simulated data. Dashed line shows the true kernel \vec{k} , which is a 32-sample function chosen to resemble the temporal impulse response of a macaque retinal ganglion cell. The LNLIF model was simulated

with parameters $g = .05$ (corresponding to a membrane time constant of 20 timesamples), $\sigma_{noise} = .5$ and $V_r = 0$. The stimulus was 30,000 time samples of white Gaussian noise with a standard deviation of .5. With only 2000 spikes of output, the estimator is able to retrieve an estimate of \vec{k} (light curve) which closely matches the true kernel, even though we only used two constraints to define each constraint set C_i in computing the likelihood. Note that the spike-triggered average (dark curve), which is an unbiased estimator for the kernel of a LNP neuron, differs significantly from this true kernel.

References

- [1] B. Aguera y Arcas and A. Fairhall. What causes a neuron to spike? <http://xxx.lanl.gov/abs/physics/0301014>, 2003.
- [2] M. Berry and M. Meister. Refractoriness and neural precision. *Journal of Neuroscience*, 18:2200–2211, 1998.
- [3] V. Bogachev. *Gaussian Measures*. AMS, New York, 1998.
- [4] E. Chichilnisky. A simple white noise analysis of neuronal light responses. *Network: Computation in Neural Systems*, 12:199–213, 2001.
- [5] A. Genz. Numerical computation of multivariate normal probabilities. *Journal of Computational and Graphical Statistics*, 1:141–149, 1992.
- [6] S. Karlin and H. Taylor. *A Second Course in Stochastic Processes*. Academic Press, New York, 1981.
- [7] J. Keat, P. Reinagel, R. Reid, and M. Meister. Predicting every spike: a model for the responses of visual neurons. *Neuron*, 30:803–817, 2001.
- [8] J. Levin and J. Miller. Broadband neural encoding in the cricket cercal sensory system enhanced by stochastic resonance. *Nature*, 1996.
- [9] L. Paninski. Convergence properties of some spike-triggered analysis techniques. *Network: Computation in Neural Systems*, Submitted, 2003.
- [10] L. Paninski, B. Lau, and A. Reyes. Noise-driven adaptation: in vitro and mathematical analysis. *Neurocomputing*, In press, 2003.
- [11] J. Pillow and E. Simoncelli. Biases in white noise analysis due to non-poisson spike generation. *Neurocomputing*, in press, 2003.
- [12] D. Reich, J. Victor, and B. Knight. The power ratio and the interval map: Spiking models and extracellular recordings. *The Journal of Neuroscience*, 18:10090–10104, 1998.
- [13] M. Rudd and L. Brown. Noise adaptation in integrate-and-fire neurons. *Neural Computation*, 9:1047–1069, 1997.
- [14] J. Victor. How the brain uses time to represent and process visual information. *Brain Research*, 886:33–46, 2000.