

Feature Selection for the Stochastic Integrate and Fire Model

Pedro Tomás, Leonel Sousa
INESC-ID / IST TULisbon
Rua Alves Redol 9, 1000-029, Lisboa – PORTUGAL
{Pedro.Tomas, Leonel.Sousa}@inesc-id.pt

Abstract – This paper presents a novel training method for estimating the parameters of integrate and fire retina models. The presented model is described by a set of linear and non-linear filters, described by basis functions and Taylor polynomials, respectively. This allows for the identification of a set of features which can be used for reproducing retina responses. A Bayesian-Laplace feature selection is proposed to choose which features can be eliminated. Thus, we are able to achieve a model using a reduced set of parameters. Experimental results show that the proposed algorithm is able to remove non-important features while still accurately reproducing retina responses.

Keywords – Integrate and Fire, Retina Modelling, Bayesian Model Selection.

I. INTRODUCTION

The modelling of the human retina has been a challenging research area and a topic of intense study in the last few years. The importance of this system is twofold. On the one hand, it allows researchers to study the connectivity of neural cells and to understand the network functionalities. The advantage of this system, regarding to the direct study of the human brain, is that it allows for an easy mapping between input stimuli and output response. On the other hand, the modelling of the human retina allows for the development of visual prostheses which could partially restore vision to blind people.

The mammal retina is composed by several layers of neurons. The network connectivity of the cells in the different layers perform a visual processing path. At the output of the retina lie the ganglion cells which transmit visual information to the brain by means of a sequence of spikes (voltage pulses).

Many different approaches have been used for modelling the response of ganglion cells to visual stimuli. Methods typically adopt spike-triggered analysis [1], information-theoretic approaches [2] or maximum-likelihood estimation [3]. One of the problems of these approaches is that they tend to require a large number of parameters, typically due to the use of a large number of coefficients for describing the linearised system. To

avoid using a large number of parameters, we have proposed the use of basis functions for describing the feedforward linear filter of integrate-and-fire (IF) models [4] (detailed in section II). Results show that the use of basis functions can significantly reduce the number of model parameters while being able to accurately model the response of the retina visual processing system. However, one of the problems when using basis functions is to know the number and the shape of the functions to apply.

In this paper we introduce a method for tuning IF models; it uses basis functions to define the linear filters. Moreover it is proposed a new method for selecting which features are important to describe the retina responses and which can be eliminated from the model. Experimental results show that the proposed algorithm is able to remove non-important features while still accurately reproducing the response of retina ganglion cells. We also show that the proposed algorithm is able to train models using non-linear functions, described by means of Taylor polynomials.

This paper is organised as follows. Section II introduces the noisy leaky IF model and section III presents an algorithm for estimating the model parameters and for performing feature selection. In section IV the proposed model and algorithm are tested by using real data from salamander retina. For assessing the performance of the training algorithm, we compare the responses from real ganglion cells with the responses of our model. Using a set of established error metrics, we also compare the results of our model with the results of the model in [4]. Finally, section VI concludes the paper.

II. INTEGRATE AND FIRE MODEL FOR THE RETINA GANGLION CELLS

Retina ganglion cells respond to visual stimuli by eliciting spikes whenever inner-membrane voltage potential surpasses a given threshold. The most complete method for representing the neuron is to use a detailed compartment model [5]; however

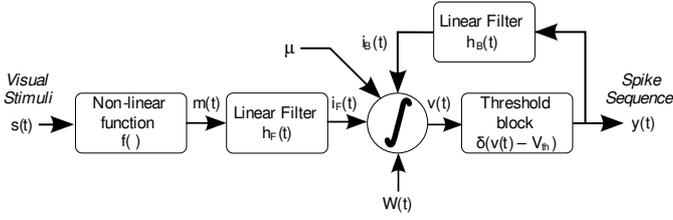


Fig. 1. GANGLION CELL RESPONSE MODEL.

this is quite complex, model which parameters are difficult to estimate. To overcome this problem, the neuron can be described by its inner-membrane voltage potential $v(t)$ which follows a stochastic differential equation (SDE) of the type:

$$dv(t) = f(v(t), t)dt + \sigma\xi(t) \quad (1)$$

where $f(v(t), t)$ denotes the deterministic dynamics of the model, $\xi(t)$ is standard white-noise and σ is a multiplicative term. A typical description of the IF neuron, assumes a two stage system. In the first stage the neuron follows a linear system with an input current $\mu + i(t)$:

$$f(V(t), t) = -\frac{1}{\tau}v(t) + i(t) + \mu \quad (2)$$

The second stage defines the instants of time when inner-membrane potential have just surpassed a threshold V_{th} . During this second stage a spike is fired by the neuron and the potential resets its value to V_0 . Under this description, the solution for (1) is:

$$v(t) = h_{IF}(t) * i(t) + \mu\tau \left(1 - e^{-t/\tau}\right) + V_0 * h_{IF}(t) + W_t \quad (3)$$

where $*$ is the convolution operator, W_t is white noise with zero mean and σ standard deviation, and $h_{IF}(t)$ is the transfer function of a low pass filter with a pole in $-1/\tau$:

$$h_{IF}(t) = \frac{1}{\tau}e^{-t/\tau}H(t) \quad (4)$$

where $H(t)$ is the Heaviside function.

The input current $i(t)$ can be defined as a sum of two components: a feedback current $i_B(t)$ that depends on the spike history and a feedforward current $i_F(t)$ (see Fig. 1). As shown in recent biological studies, these two components are non-linear functions with temporal dynamics. However, in this work, we apply only a non-linearity to the input visual stimuli. Experimental results presented in section IV show that this model is able to accurately reproduce the response of ganglion cells for the testing data set.

The input component is therefore describe by a two stage model. First, the input stimuli is transformed by means of a non-linear function $m(t) = f(s(t))$. Secondly, the output is convoluted with a linear filter with impulse response $h_F(t)$. The feedback component is described by a linear filter with impulse response $h_B(t)$. Finally, the spike response is obtained by using

a threshold block, such that a spike is fired whenever the inner-membrane potential $v(t)$ surpasses V_{th} . For tuning the model, the following discretisation was applied:

$$i[n] = \overbrace{h_F[n] * f(s[n])}^{i_F[n]} + \overbrace{h_B[n] * y[n]}^{i_B[n]} \quad (5a)$$

$$v[n] = \beta \cdot v[n-1] + (1 - \beta) \cdot (\mu + i[n] + W_n) \quad (5b)$$

$$y[n] = H(v[n] - V_{th}) \quad (5c)$$

where the non-linear function is described by Taylor polynomials. The filters $h_F[n]$ and $h_B[n]$ are constructed using orthogonal Laguerre basis functions [6] with z -domain transfer function:

$$h_L^{(k)}[z] = \frac{\sqrt{1-|\epsilon|}}{1-\epsilon z} \prod_{i=1}^{k-1} \frac{z-\epsilon}{1-\epsilon z} \quad k = 1, \dots, M \quad (6)$$

where ϵ regulates the position of the basis poles and zeros.

Thus, the input currents $i_F(t)$ and $i_B(t)$ are described as

$$i_F[n] = \sum_{i=1}^{N_F} \sum_{j=1}^{N_P} a_i b_j \left(h_L^{(i)} * s^j \right) [n] = \mathbf{A}^T (\mathbf{\Lambda}_F[n]) \mathbf{B} \quad (7a)$$

$$i_B[n] = \sum_{i=1}^{N_B} c_i \left(h_L^{(i)} * y \right) [n] = (\mathbf{\Lambda}_B[n])^T \mathbf{C} \quad (7b)$$

where $\mathbf{A} = [a_k]$, $\mathbf{B} = [b_k]$, $\mathbf{C} = [c_k]$ are column vectors of size N_F , N_P and N_B , respectively, and

$$(\mathbf{\Lambda}_F[n])_{ij} = (h_{IF} * h_L^{(i)} * s^j) [n] \quad (8a)$$

$$(\mathbf{\Lambda}_B[n])_i = (h_{IF} * h_L^{(i)} * y) [n] \quad (8b)$$

Since the model is affected by white noise, v_n follows a normal distribution with mean \bar{v}_n and standard deviation σ_n . Thus, it results:

$$\bar{v}_n = \mu\Lambda_\mu[n] + \mathbf{A}^T (\mathbf{\Lambda}_F[n]) \mathbf{B} + (\mathbf{\Lambda}_B[n])^T \mathbf{C} \quad (9a)$$

$$\Lambda_\mu[n] = 1 - \beta^n \quad (9b)$$

$$\sigma_n^2 = \sigma^2 (1 - \beta)^2 \frac{1 - \beta^{2n}}{1 - \beta^2} + \beta^{2n} \sigma_0^2 \quad (9c)$$

where σ_0 is the noise standard deviation after a spike has been fired, and $\Lambda_\mu[n]$ and σ_n are obtained by noting that the filter h_{IF} establishes a geometric series. Considering that during the cell discharge stage (period after a spike has been fired) the cell is affected by noise with the same properties as in the remaining period, we can replace σ_n by σ_∞ (corresponding to the noise level when $n \rightarrow \infty$):

$$\sigma_\infty = \sigma \frac{1 - \beta}{\sqrt{1 - \beta^2}} \quad (10)$$

Accordingly to the previous description, the complete set of parameters for this model would be:

$$\Theta = \{ \mathbf{A}, \mathbf{B}, \mathbf{C}, \mu, \beta, V_0, V_{th}, \sigma \} \quad (11)$$

However, careful analysis of the system can help eliminate some of the model parameters. The set $\{V_0, V_{th}\}$ is dependent on the gains a_k, b_k, c_k, μ : in order to keep the same model response, to increase the difference $V_{th} - V_0$ implies increasing the gains of the system parameters; similarly, adding an offset to both V_{th} and V_0 implies a change in the constant input current μ . Thus we have set $V_0 = 0$ and $V_{th} = 1$.

The parameter β affects the behaviour of the integrator block in Fig. 1. However, a careful examination of the model can help to exclude β from the set of tunable parameters. Knowing that the convolution of two linear filters is a linear filter, one can simply fix the value for β and let the system temporal transfer function be defined by the filters h_F and h_B . Considering a sample time T , one can set β such that the module of the system pole (defined by $s_{pole} = \frac{1}{T} \log \beta$) is sufficiently high; the influence of known undesirable frequencies is reduced while the influence of unknown frequencies is defined by the coefficients of a_k and b_k of the filters h_F and h_B , respectively. In our experiments it was set $\beta = 0.9$ for a sampling time $T = 0.001$ s. The tunable parameter set is therefore reduced from Θ to Θ^* :

$$\Theta^* = \{\mathbf{A}, \mathbf{B}, \mathbf{C}, \mu, \sigma\} \quad (12)$$

The above analysis on the parameter set has intentionally left out four parameters: ϵ, N_F, N_B and N_P . The first parameter, which affects the shape of the basis functions, is difficult to optimise since it introduces local minimums in the optimisation function (presented further on). A simple bypass to this problem is to introduce a sufficiently large number of basis functions to allow a good representation of real transfer function of the retina cells. However adding too many basis functions leads to a model which learns to represent the desired output (presented during the training step) and not the retina transfer function. Similar constraints could be argued against the degree of Taylor polynomials. The described problem is one of the main focus of this paper, addressed in subsection III.A.

III. MODEL OPTIMISATION

Following a Bayesian approach, one can define the probability of the output $y_1, \dots, y_n = \mathbf{y}_n, y_n \equiv y[n]$, as

$$P(\mathbf{y}_n | \mathbf{s}_n) = \prod_{n=1}^M P(y_n | \mathbf{y}_{n-1}, \mathbf{s}_n) \quad (13)$$

where $\mathbf{s}_n = s_1, \dots, s_n, s_n \equiv s[n]$, represents the visual stimuli. Also $P(y_n | \dots)$ represents the probability for the estimated output \hat{y}_n to be equal to the real output y_n . We assume $y_n = 1$ if a spike has been fired at time step n and $y_n = 0$ otherwise.

The presented model is a hidden Markov model (HMM) where the subthreshold potential v_n is unknown. Thus, one must marginalise the probabilities $P(y_n | \mathbf{y}_{n-1})$:

$$\begin{aligned} P(y_n | \mathbf{y}_{n-1}, \mathbf{s}_n) &= \int_{-\infty}^{+\infty} P(y_n | v_n) p(v_n | \mathbf{y}_{n-1}, \mathbf{s}_n) \\ &= y_n + (-1)^{y_n} N_{cdf}(V_{th} | \mathbf{y}_{n-1}, \mathbf{s}_n) \end{aligned} \quad (14)$$

where $N_{cdf}(V_{th} | \bar{v}_n, \sigma_\infty)$ represents the normal cumulative distribution function of $v_n \sim N(\bar{v}_n, \sigma_\infty)$ evaluated at V_{th} .

For model tuning, the parameter set Θ that maximises the probability of the sequence must be chosen (or equivalently, the log-probability). Thus, differentiating the log-probability $\log P(\mathbf{y}_n | \mathbf{s}_n)$ in order to the parameters results:

$$\begin{aligned} \nabla l(\Theta) &= \sum_{n=1}^M \frac{\nabla P(y_n | \mathbf{y}_{n-1}, \mathbf{s}_n)}{P(y_n | \mathbf{y}_{n-1}, \mathbf{s}_n)} = \\ &= \sum_{n=1}^M \frac{-(-1)^{y_n} p_{th}}{P(y_n | \mathbf{y}_{n-1}, \mathbf{s}_n)} \left[\nabla \bar{v}_n + \frac{V_{th} - \bar{v}_n}{\sigma_\infty} \nabla \sigma_\infty \right] \end{aligned} \quad (15)$$

where $p_{th} = p(v_n = V_{th} | \mathbf{y}_{n-1}, \mathbf{s}_n)$ represent the probability density function of v_n evaluated at V_{th} .

Notice that the optimisation function in (15) formally coincides with the first passage time (FPT) method [3].

A. Applying Bayesian feature selection

So far we have described the IF model used for estimating the output of retina ganglion cells. Additionally we have described an optimisation function for tuning the model parameters. However, as previously referred, we have neither discussed how many basis functions should one use for modelling the retina processing system, nor the shape of these functions (parameter ϵ in (6)). Since ϵ is not easily optimisable, a better way for tuning the model would be to generate a set of basis functions and then to select the minimum set of functions which best describe the ganglion cells' transfer function. While such a method has been an intense topic of research for application in several areas, it has never been applied for feature selection in IF models.

Suppose that we have a set of candidate models \mathcal{M} each described by set of parameters $\Theta_m = \{\theta_1, \dots, \theta_{K_m}\}$ (not necessarily of equal size). Using a Bayesian procedure, the best model for the system would be the one which maximises the joint probability of the output sequence \mathbf{y} and the model m :

$$p(\mathbf{y}, m | \mathbf{s}) = p(m) \int_{\Theta_m} p(\mathbf{y} | \mathbf{s}, \Theta_m, m) p(\Theta_m | m) d\Theta \quad (16)$$

Since the integral is not always easy to compute one can use the large-scale Bayesian-Laplace approximation [7]:

$$\begin{aligned} (\hat{m}, \hat{\Theta}) &= \arg_{(m, \Theta)} \max p(\mathbf{y}, m | \mathbf{s}) \\ &\approx \arg_{(m, \Theta)} \left\{ \log p(\mathbf{y} | \mathbf{s}, \hat{\Theta}_m, m) + \frac{d}{2} \log(2\pi) + \right. \\ &\quad \left. - \frac{1}{2} \log \det \mathbf{I}_F(\mathbf{y} : \hat{\Theta} | m) \right\} \end{aligned} \quad (17)$$

where $\mathbf{I}_F(\mathbf{y} : \hat{\Theta} | m)$ is the observed Fisher information matrix:

$$\mathbf{I}_F(\mathbf{y} : \hat{\Theta} | m) = \mathbb{E} \left[-\frac{\partial^2}{\partial \theta_i \partial \theta_j} \log p(\mathbf{y} | \mathbf{s}) \right]_{\Theta = \hat{\Theta}} \quad (18)$$

Algorithm 1 Model optimisation algorithm

```

< INITIALISE MODEL  $m^*$  >
Set:  $V_{th} \leftarrow 1, \beta \leftarrow 0.9$ 
Generate the set of basis functions and set the initial non-linearity order:
 $N_F \leftarrow 20, N_B \leftarrow 20, N_P \leftarrow 10$ 
Set:  $a_k \leftarrow 0, b_k \leftarrow 0, c_k \leftarrow 0, b_1 \leftarrow 1, \mu \leftarrow 0, \sigma \leftarrow 0.2$ .
Set:  $\Theta^* \leftarrow \arg_{\Theta} \max \log p(\mathbf{y}|\mathbf{s}, \Theta, m^*)$ 
Set:  $L^* \leftarrow \log p(\mathbf{y}|\mathbf{s}, \Theta^*, m^*)$ 

< MODEL OPTIMISATION >
repeat
  Set IsBetter  $\leftarrow$  false

  < GENERATE NEW MODEL BY CHANGING  $N_F$  >
   $\hat{m} \leftarrow m^*, N_F(\hat{m}) \leftarrow N_F(m^*) \pm 1$ 
   $(\Theta^*, L^*, m^*, \text{IsBetter}) \leftarrow \text{optimise}(\Theta^*, \hat{m}, m^*, L^*, \text{IsBetter})$ 

  < GENERATE NEW MODEL BY CHANGING  $N_B$  >
   $\hat{m} \leftarrow m^*, N_B(\hat{m}) \leftarrow N_B(m^*) \pm 1$ 
   $(\Theta^*, L^*, m^*, \text{IsBetter}) \leftarrow \text{optimise}(\Theta^*, \hat{m}, m^*, L^*, \text{IsBetter})$ 

  < GENERATE NEW MODEL BY CHANGING  $N_P$  >
   $\hat{m} \leftarrow m^*, N_P(\hat{m}) \leftarrow N_P(m^*) \pm 1$ 
   $(\Theta^*, L^*, m^*, \text{IsBetter}) \leftarrow \text{optimise}(\Theta^*, \hat{m}, m^*, L^*, \text{IsBetter})$ 

until IsBetter=false

< AUXILIARY OPTIMISATION FUNCTION >
function  $(\Theta^*, L^*, m^*, \text{IsBetter}) = \text{optimise}(\Theta^*, m, m^*, L^*, \text{IsBetter})$ 
  Set:  $\hat{\Theta} \leftarrow \arg_{\Theta} \max \log p(\mathbf{y}|\mathbf{s}, \Theta, m)$ 
  Set:  $\hat{L} \leftarrow \log p(\mathbf{y}|\mathbf{s}, \hat{\Theta}, m)$ 
  if  $\log p(\mathbf{y}|\mathbf{s}, \hat{\Theta}, m) > \log p(\mathbf{y}|\mathbf{s}, \Theta^*, m^*)$  then
    Set:  $\Theta^* \leftarrow \hat{\Theta}$ 
    Set:  $L^* \leftarrow \hat{L}$ 
    Set:  $m^* \leftarrow m$ 
    Set IsBetter  $\leftarrow$  true
  end if
end function

```

The above equation assumes that: *i*) there is no prior knowledge on the best model for our data, and *ii*) that $p(\Theta_m|m)$ is sufficiently flat around $\hat{\Theta}$.

While it is possible to compute the Fisher information matrix for the presented model, one can use large-scale asymptotics [7]. This results in a new maximisation function L :

$$L(\Theta_m; m) = \log p(\mathbf{y}|\mathbf{s}, \Theta_m, m) - \frac{d}{2} \log N \quad (19)$$

where $d = N_F + N_B + N_P + 2$ is the dimension of Θ_m and N is the number of valid output samples, i.e., is the number of samples for which there is knowledge on both the stimuli input and the spike history.

Experimental results have shown that the asymptotics case leads to a faster algorithm (since it is not required to compute the observed Fisher information matrix) and that the results are not influenced by the large-scale approximation (for a review on other model selection techniques see [8]).

B. Optimisation algorithm

One of the problems in feature selection is the lack of a differentiable function on the model (or, in our specific case, on the number of features). As presented in Algorithm 1, a method to overcome this difficulty is to iteratively add/remove

features from the current solution and then to maximise $p(\mathbf{y}|\mathbf{s})$ in order to the current set of parameters. This method works as follows. First an initial model m^* is created using a large number of basis functions $N_F = N_B = 20$ and a high order Taylor polynomial $N_P = 10$. This model is initialised by setting all free parameters to zero with the exception of the coefficient for the first order polynomial of the Taylor expansion series ($b_1 = 1$). Additionally, the noise standard deviation σ must be initialised such that $\sigma > \sqrt{1 - \beta^2}/4$ to avoid numerical representation problems.

During the first phase of the algorithm, the initial model m^* is tuned by using (15) to find the parameters Θ^* that maximise (13). Adaptive steps [9] are used to speed up the convergence of the gradient ascent process.

After the initial parameter tuning, the algorithm iteratively tries to find new models \hat{m} such that $L(\hat{\Theta}, \hat{m}) > L(\Theta^*, m^*)$. Each new model \hat{m} is created from the *old best* model m^* and then modified by increasing/decreasing the number of basis functions or by increasing/decreasing the order of the non-linear function. To search for new models \hat{m} , the algorithm uses the following generation order:

1. $N_F(\hat{m}) = N_F(m^*) - 1$
2. $N_F(\hat{m}) = N_F(m^*) + 1$
3. $N_B(\hat{m}) = N_B(m^*) - 1$
4. $N_B(\hat{m}) = N_B(m^*) + 1$
5. $N_P(\hat{m}) = N_P(m^*) - 1$
6. $N_P(\hat{m}) = N_P(m^*) + 1$

For each step 1-6, once a new model has been found such that $L(\hat{\Theta}, \hat{m}) > L(\Theta^*, m^*)$, the algorithm proceeds to the next step. Thus, for example, after removing one forward basis function it only tries to remove another after going through steps 2-6.

Moreover, when removing one basis function (steps 1 and 3) one should first try to remove the basis having the least influence on the model. To assess basis influence the algorithm computes their total power ψ_k :

$$\begin{cases} \psi_k = |a_k| \sqrt{T \sum_{n=0}^{\text{basis length}} (h_L^{(k)}[n])^2} & , \text{ for feedforward basis} \\ \psi_k = |c_k| \sqrt{T \sum_{n=0}^{\text{basis length}} (h_L^{(k)}[n])^2} & , \text{ for feedback basis} \end{cases} \quad (20)$$

and then starts by removing the function having the lowest power.

IV. EXPERIMENTAL RESULTS

The proposed training algorithm was implemented and tested with the experimental data used in [10]. These data consists of 12 trials of full field white noise stimulation for a salamander ON cell, where each trial has a duration of 10 seconds with an average count of 8.34 spikes per second. The visual stimuli was normalised by subtracting its mean value and then diving by its standard deviation. The resulting stimuli, which corresponds

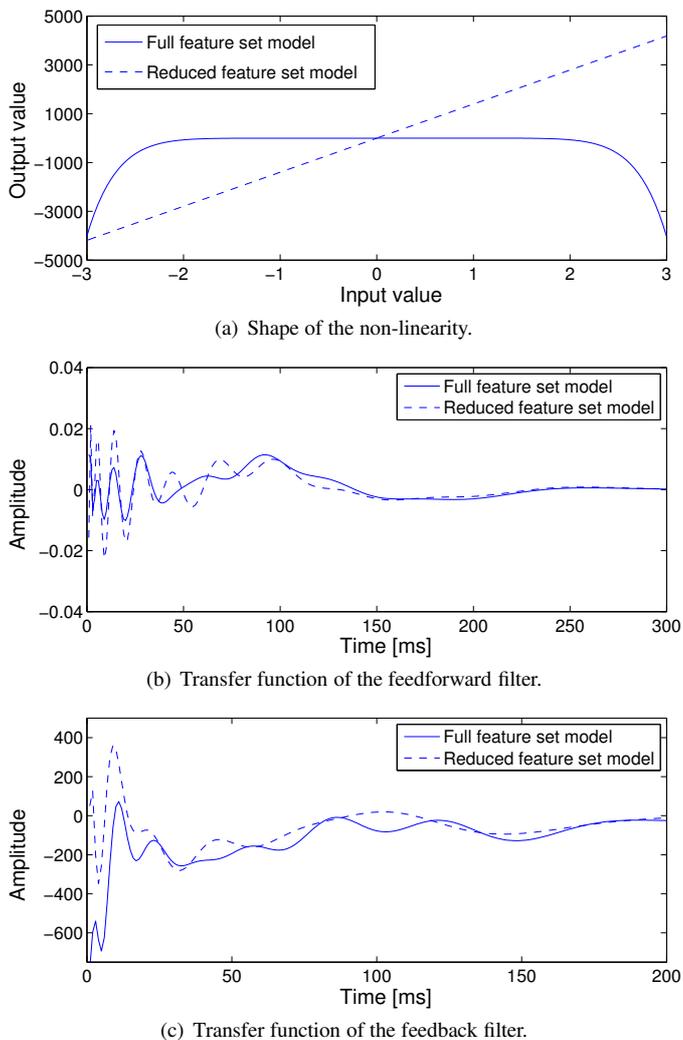


Fig. 2. MODEL FUNCTIONALS WHEN USING THE COMPLETE SET OF BASIS (WITHOUT FEATURE SELECTION) AND WHEN USING A REDUCED SET OF BASIS (WITH FEATURE SELECTION).

to the input s_n given to the model, is therefore a sequence of normally distributed random values with zero mean and unitary standard deviation.

This data was used for tuning the parameters of several IF models. The differences between the experimented models concerned the selected basis functions and the leaky integrator coefficient β . The values for the experimented parameters ranged in the intervals $0.6 \leq \epsilon \leq 0.9$ and $0.7 \leq \beta \leq 0.95$. Experimental tests ranged from using orthogonal and non-orthogonal basis (by using basis extracted with different values of ϵ). Our tests revealed little differences in the model response, which corresponded to small differences on the error measures presented below. For evaluating the performance of the feature selection algorithm, results are presented for only one tuned model. A set of 20 basis functions are used for both the feedforward and feedback filters ($N_F = N_B = 20$). The non-

linearity was initially set to $N_P = 10$. Fig. 2 presents the results of the model estimation with the feature selection mechanism, “*Reduced feature set model*”, and without it, “*Full feature set model*”. These results show that the feature selection mechanism was able to reduce the number of parameters from an initial value of 52 to only 18 (resulting in $N_F = 10$, $N_B = 5$, $N_P = 1$).

For the used testing data, the proposed training algorithm was able to remove the non-linearity all together. However, this might not be the case when modelling the responses of other ganglion cells: it is widely known that the processing mechanism in the retina includes several paths, some of which having linear behaviour, while have non-linear and dynamic behaviours. Thus, the possibility of including non-linearities in the model is an important issue and, as far as we are aware, this is the first time a non-linear IF model is presented and tested on real data.

The shape of the linear and non-linear filters for both the “*Reduced feature set model*” and the “*Full feature set model*” cases are presented in Fig. 2. After model tuning, 100 spike sequence trials were generated for each of the two models. The first 12 trials for each model are presented in Fig. 3. The real responses from the retina ganglion cells are also presented in this figure.

For comparing the response of the models to visual stimuli against real data, two error metrics proposed in [11] were used. The first metric accounts for the cost associated with the absolute time of occurrence of neuronal events (Spike Time Metric). The second metric accounts for the cost of changing the intervals between two spikes (Inter Spike Metric). The cost of moving a spike was set to $q = 50 \text{ s}^{-1}$ (see [11]). Obtained values for the error metrics using the 12 trials from real data and the 100 trials from our models are shown in Table I. From the values presented in the table, it can be concluded that the proposed algorithm for tuning the parameters of the IF model is able to reduce the number of model components (or free parameters) without significantly affecting the performance of the model.

We also compare the obtained results with a previously proposed IF model using basis functions [4]. In this model only the feedforward filter was modelled by basis functions. The feedback filter is represented by an exponentially decaying function and no non-linearity block was included. The results for the Spike Time and Inter Spike error metrics are also presented in Table I. As shown, the errors for both metrics are smaller for the proposed model (with and without feature selection) than those for the exponentially decaying model previously proposed.

A firing rate metric was also used as an auxiliary error measure, namely the normalised mean squared error (NMSE) [12] was applied. The firing rates were estimated for both the real and the estimated data, by convolving their PeriStimulus Time Histogram (PSTH) [1] with a Gaussian window of zero mean and 20ms of standard deviation. Using this metric it can also be observed that the difference between the full and the reduced parameters set models is very small. Once again, it is observed that the proposed model presents a smaller NMSE error than the exponentially decaying model [4].

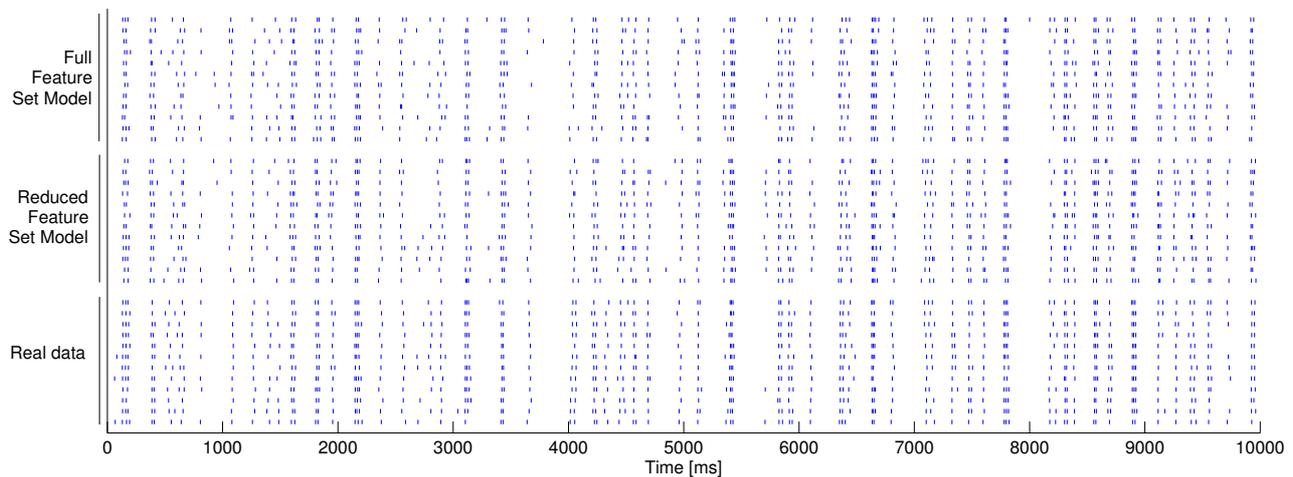


Fig. 3. 12 RESPONSE TRIALS FOR BOTH REAL RETINA GANGLION CELLS AND THE PRESENTED MODELS.

TABLE I.
ERROR MEASURES BETWEEN TRAINED MODELS AND REAL RESPONSES.

		Spike Time	Inter Spike	Spike Count	NMSE
<i>Retina ganglion cells response</i>					
Real Data	mean	42.19	59.05	83.42	
	std	3.58	5.26	2.81	
<i>Full feature set model: $N_F = 20, N_B = 20, N_P = 10$</i>					
I&F	mean	56.88	71.98	83.04	
	std	4.43	4.84	2.72	
Real vs I&F	mean	58.31	74.59		0.101
	std	3.70	5.07		
<i>Reduced feature set model: $N_F = 10, N_B = 5, N_P = 1$</i>					
I&F	mean	57.99	72.93	83.31	
	std	4.44	5.16	2.90	
Real vs I&F	mean	57.71	73.12		0.093
	std	4.38	4.96		
<i>IF model with exponentially decaying feedback filter [4]</i>					
I&F	mean	62.17	76.92	83.08	
	std	4.15	3.98	2.50	
Real vs I&F	mean	63.28	81.09		0.249
	std	3.71	4.36		

mean - mean result std - standard deviation

V. ACKNOWLEDGEMENTS

The experimental data used in this paper was gently provided by Prof. Markus Meister.

VI. CONCLUSIONS

This paper presents a non-linear noisy-IF model for the modelling of the retina ganglion cells to visual stimuli. Carefull analysis of the presented model allows to identify the parameters that do not influence the training. Furthermore, a novel training algorithm is presented which includes a feature selection mechanism. This algorithm is able to automatically decide on

the number of basis functions and on the order of the non-linearity used for modelling the real data. Experimental results show that the proposed algorithm was able to significantly reduce the number of features without compromising the model performance. They also show that the proposed model achieves lower error measures than the previously proposed IF model using basis function.

REFERENCES

- [1] E. J. Chichilnisky, "A simple white noise analysis of neuronal light responses," *Network: Computation in Neural Systems*, vol. 12, no. 2, pp. 199–213, May 2001.
- [2] J. W. Pillow and E. P. Simoncelli, "Dimensionality reduction in neural models: an information-theoretic generalization of spike-triggered average and covariance analysis," *Journal of Vision*, vol. 6, no. 4, pp. 414–428, April 2006.
- [3] Liam Paninski, "The most likely voltage path and large deviations approximations for integrate-and-fire neurons," *Journal of Computational Neuroscience*, vol. 21, no. 1, pp. 71–87, April 2006.
- [4] S. Capela, P. Tomás, and L. Sousa, "Stochastic integrate-and-fire model for the retina," in *15th European Signal Processing Conference (EUSIPCO'2007)*, September 2007.
- [5] Andreas V. M. Herz, Tim Gollisch, Christian K. Machens, and Dieter Jaeger, "Modeling Single -Neuron Dynamics and Computations: Balance of Detail and Abstraction," *Science*, vol. 314, pp. 80–85, October 2006.
- [6] H. Akçay and B. Ninness, "Orthonormal basis functions for modelling continuous-time systems," *Signal Processing*, vol. 77, no. 1, pp. 261–274, 1999.
- [7] J. Oliver and R. Baxter, "MML and Bayesianism: Similarities and differences (introduction to Minimum Encoding Inference - part II)," Tech. Rep. 206, Department of Computer Science, Monash University, 1994.
- [8] A. D. Lanterman, "Schwarz, Wallace, and Rissanen: intertwining themes in theories of model selection," *International Statistical Review*, vol. 69, no. 2, pp. 185–212, 2001.
- [9] L. B. Almeida, T. Langlois, J. D. Amaral, and A. Plakhov, *Parameter Adaptation in Stochastic Optimization*, chapter 6, Cambridge University Press, 1998.
- [10] J. Keat, P. Reinagel, R. C. Reid, and M. Meister, "Predicting every spike: A model for the responses of visual neurons," *Neuron*, vol. 30, pp. 803–817, June 2001.
- [11] Jonathan D. Victor and Keith P. Purpura, "Metric-space analysis of spike trains: theory, algorithms and application," *Computation Neural Systems*, vol. 8, pp. 127–164, 1997.
- [12] M. J. Berry II and M. Meister, "Refractoriness and neural precision," *The Journal of Neuroscience*, 1998.