Modeling gain control and inferring gain from data using GLMs

Eugenio Piasini

January 21, 2021

Abstract

Here I develop a very simple model of a neuron with gain control, and I show that it is possible to use a GLM to extract a *gain modulation index* that captures this phenomenon.

1 From gain to data: forward model and simulation

In our model, time evolves in discrete steps. We model the sensory stimulus as a simple scalar signal

$$x_t \sim \mathcal{N}(\mu, \sigma_t)$$

where $t \in \mathbb{N}$ is the time, μ a constant, and we call σ_t the *contrast* at time t. With respect to Chris's experiment, this is like considering a stimulus with only one frequency.

The model neuron has a firing rate that at time t depends only on the stimulus x_t and the contrast σ_t at that same time. This is another simplification with respect to how Chris analyses his data, as we ignore any time-delayed dependence of the neural response on previous stimuli. More specifically, we assume that the number of spikes y_t emitted by the neuron at time t follow a Poisson probability distribution:

$$y_t \sim \text{Poisson}(\lambda_t) \tag{1}$$

where λ_t is the firing rate at time t, given by

$$\lambda_t = \exp\left[a + g(\sigma_t)b(x_t - c)\right] \tag{2}$$

where $g(\cdot)$ is a *gain control function*, and a, b and c are parameters of the model. The parameter c represents the *operating point of the gain*. In absence of gain modulation (if g was constant), c and a could be varied together in a way that does not affect the behavior of the model, but when g depends nontrivially on σ this degree of freedom is removed, and the specific choice of c and a matters. Another similar degree of freedom is that only the product of g and b matters; for ease of interpretation we will fix this by requiring g to be adimensional and such that

$$\frac{1}{2}\left[g(\sigma_H)+g(\sigma_L)\right]=1 \tag{3}$$

where σ_H and σ_L are the "high" and "low" contrast values. The idea behind this constraint is that we want to set the neutral value $g=1$ to be the midpoint between the gain in the high and low condition (we will work out how this constrains g in practice in the next section).

Finally, note also that we assume that here the contrast at time t is known and does not need to be estimated from the stimulus. When multiple frequency bands are available, the contrast can simply be estimated as the standard deviation of the input across frequency bands.

1.1 Optimal gain control

In the spirit of the efficient coding principle, we can easily derive a form for $g(\sigma)$ that will guarantee that, under certain conditions, the dynamic range of the neuron will be approximately conserved under changes in contrast. To do this, define the dynamic range as

$$\begin{aligned} R(\sigma) &= \lambda(\mu+\sigma)-\lambda(\mu-\sigma) \\ &= e^a\left[\exp\left(g(\sigma)b(\mu+\sigma-c)\right)-\exp\left(g(\sigma)b(\mu-\sigma-c)\right)\right] \end{aligned}$$

If the argument of the exponentials is not too large, we can linearize this expression to obtain

$$R(\sigma)\simeq 2e^a b\sigma\cdot g(\sigma) \tag{4}$$

and we see that R is approximately independent of σ provided that $g(\sigma)\propto 1/\sigma$. For our model, we will set

$$g(\sigma)=\frac{\bar{\sigma}}{\sigma} \tag{5}$$

Where $\bar{\sigma}$ is the harmonic mean of σ_H and σ_L:

$$\bar{\sigma}:=\left[\frac{1}{2}\left(\frac{1}{\sigma_H}+\frac{1}{\sigma_L}\right)\right]^{-1}=2\frac{\sigma_H\sigma_L}{\sigma_H+\sigma_L} \tag{6}$$

This choice is justified by verifying that Equation 5 satisfies the requirements for g laid out in the previous section. Indeed, with this definition g is clearly adimensional, and moreover

$$g(\sigma_H)+g(\sigma_L)=\frac{\bar{\sigma}}{\sigma_H}+\frac{\bar{\sigma}}{\sigma_L}=\bar{\sigma}\left(\frac{\sigma_L+\sigma_H}{\sigma_H\sigma_L}\right)=2$$

so that Equation 3 holds.

Finally, for the purpose of illustrating the effect of gain control in the simulations, we will also consider an interpolation scheme that smoothly transforms a model with gain control to a similar model without gain control. To do this, we will substitute g in Equation 2 as follows:

$$g(\sigma) \longrightarrow \xi g(\sigma) + (1-\xi) \quad , \quad 0 \leq \xi \leq 1 \tag{7}$$

so that by changing ξ we can decide whether gain control should be at the optimal level ($\xi = 1$), off entirely ($\xi = 0$) or something in between.

Putting everything together, the final expression for the firing rate of the forward model is

$$\lambda_t = \exp\left[a + b\left(\xi \frac{\bar{\sigma}}{\sigma_t} + (1-\xi)\right)(x_t - c)\right] \tag{8}$$

1.2 Simulation

We will simulate the neuron for N trials, each of which will be composed of $2T$ timepoints; σ_t will change during the trial according to

$$\sigma_t = \begin{cases} \sigma_L & \text{if } t < T \\ \sigma_H & \text{if } t \geq T \end{cases}$$

with $\sigma_L < \sigma_H$ (this represents a "low contrast to high contrast" experiment).

2 From data to gain: Generalized Linear Model

The forward model developed in the previous sections seems like a reasonable way to think about gain control in its simplest possible terms. It also has another advantage: the model, as described by Equation 1 and Equation 8, can in principle be represented exactly by a Poisson GLM, provided that the predictors are chosen appropriately. This means that we can generate data from the forward model, fit the data with a GLM, and hope to recover the essential patterns of interest (most importantly, gain control). And as a consequence of this last point, as long as we believe that the forward model is a reasonable description of what the experimentally-recorded neurons are doing, then we can be relatively confident that defining a *gain modulation index* based on a GLM fit is not a terrible idea.

2.1 Definition of the GLM

The inference model is a Poisson GLM with an intercept term and the following predictors:

$$(x_t - \mu) \quad , \quad \frac{\bar{\sigma}}{\sigma_t} \quad , \quad (x_t - \mu)\frac{\bar{\sigma}}{\sigma_t}$$

In other words, we have a stimulus predictor, a contrast predictor, and their interaction. Note that there could be many other combinations of predictors that would yield an equivalent GLM (for instance the predictors can be freely rescaled, as they are always z-scored prior to fitting anyway, so there is clearly no need for the $\bar{\sigma}$ term), but this particular choice makes the equations prettier and helps in drawing connections with the parameters that define the forward model.

The GLM therefore models the data at time t as a Poisson distribution with the following mean:

$$\lambda_t = \exp\left[\beta_0 + \beta_1(x_t - \mu) + \beta_2\frac{\bar{\sigma}}{\sigma_t}(x_t - \mu) + \beta_3\frac{\bar{\sigma}}{\sigma_t}\right] \tag{9}$$

where $\beta_0 \ldots \beta_3$ are the parameters to be inferred.

2.2 Gain modulation index

By comparing Equation 9 to Equation 8 we immediately see that the two expressions, seen as functions of x_t and σ_t can be made equivalent with an appropriate choice of the β parameters. At the moment we're not too interested in deriving this full correspondence, but we observe that Equation 9 can be rewritten as follows

$$\begin{aligned}
\log(\lambda_t) &= \beta_0 + \beta_1(x_t - \mu) + \beta_2\frac{\bar{\sigma}}{\sigma_t}(x_t - \mu) + \beta_3\frac{\bar{\sigma}}{\sigma_t} \\
&= \beta_0 - (\beta_1 + \beta_2)\mu + \beta_3\frac{\bar{\sigma}}{\sigma_t} + \left(\beta_1 + \beta_2\frac{\bar{\sigma}}{\sigma_t}\right)x_t \\
&= \beta_0 - (\beta_1 + \beta_2)\mu + \beta_3\frac{\bar{\sigma}}{\sigma_t} + \left(\beta_1 + \beta_2 + \beta_2\frac{\bar{\sigma}}{\sigma_t} - \beta_2\right)x_t \\
&= \beta_0 - (\beta_1 + \beta_2)\mu + \beta_3\frac{\bar{\sigma}}{\sigma_t} + (\beta_1 + \beta_2)\left[1 + \frac{\beta_2}{\beta_1 + \beta_2}\left(\frac{\bar{\sigma}}{\sigma_t} - 1\right)\right]x_t
\end{aligned} \tag{10}$$

We see then that the quantity

$$w(\sigma) := 1 + \frac{\beta_2}{\beta_1 + \beta_2}\left(\frac{\bar{\sigma}}{\sigma_t} - 1\right) \tag{11}$$

corresponds to g in the forward model. Indeed, in the expression for $\log(\lambda)$, in both models x_t is multiplied by a dimensional parameter independent from the gain (b in Equation 2, $\beta_1 + \beta_2$ in Equation 10) and by something that depends on the contrast, that is adimensional, and that is 1 when the contrast matches some reference level (chosen "by hand" to be $\bar{\sigma}$ in both cases). This is g in Equation 8 and w in Equation 10. Therefore, it is possible to interpret w as a *gain modulation index* obtained from the GLM fit.

2.3 Fitting procedure and regularization

While not strictly necessary for the purposes of this simulation, in view of the application of this method to real data we will fit the GLM using regularization. In my experience using elastic net (I tried $\alpha = 0.95$ in `glmnet`, which is a personal favorite) yields the cleanest results. On the other hand, the results one gets using ridge regression ($\alpha = 0$) are close enough, which is important because ridge is much cheaper computationally than elastic net, and this would matter when running a version of this analysis on the very large amount of data collected by Chris. The main difference I have found is that in the case of a simulation with $\xi = 0$ (namely, no gain control at all), elastic net correctly infers β_2 to be *exactly* 0, leading to $w(\sigma) \equiv 1$. With ridge regression, β_2 is typically inferred as a very small number, which results in w being weakly dependent on σ but not exactly constant. In any case, qualitatively the result is correct: w is much closer to being constant for data simulated with $\xi = 0$ than for data simulated with positive values of ξ. Moreover, this is just a fine point that is discernible because we control the forward model and because the GLM is perfectly specified with respect to that — in the real world, other sources of noise and misspecification are likely to be more important than the difference between regularization methods.

Another thing to note regarding the regularization is that I have observed the fit quality to degrade when the forward model is only weakly responsive to the input — i.e. when b is small[1]. This happens because the regularization (in both the ridge and elastic net case) biases the inference procedure towards fitted parameters that are small in magnitude. So when overall the observed spike counts do not seem to change much over the course of the simulation, the regularized fit will tend to only look at the average firing rate, and shrink away the other coefficients (if β_1, β_2 and β_3 are small, then the average firing rate is set by β_0).

2.4 Interpretation and possible weaknesses

In the previous sections, we have set up the forward model and the GLM in such way that in principle a perfect GLM fit should be able to recover not only the general patterns in the data, but also the actual value of the parameters of the forward model. For instance, we have seen that in a perfect fit we would have $\beta_1 + \beta_2 = b$ and $w = g$. This provides us with a way to sanity-check the fit, gauging how well the fit is performing in an absolute sense. For instance, empirically I find that $\beta_1 + \beta_2$ tends to be close, but slightly smaller, than the value I use for b in the forward model (if $b = 0.1$, I may find that $\beta_1 + \beta_2 = 0.099$ — of course the specific numerical values depend on the other quantities involved in the simulation). This is due to the regularization which, as mentioned in the previous section, tends to shrink the values of the coefficients towards zero.

[1] Initially, I observed this when playing around with g and trying to multiply it by some constant, which is the same as scaling b. Chris, this is the issue we discussed at some point during the call on Jan 20th.

Parameter	Value
μ	30
σ_H	5
σ_L	2
N. of timesteps	2×20
N of trials	500
a	log(50)
b	0.1
c	30 ($= \mu$)
α	0.95

Table 1: Simulation parameters for the figures in this document.

In any case, all of this is only useful as a way of justifying the approach I propose here, and to show that it is at least consistent in a case where we can set ourselves up with the best chances of success by ensuring that the data is generated by a process that can be fully captured by our inference model. When one moves to real data, one needs to trust that the quantities of interest ($\beta_1 + \beta_2$, w) are meaningful, and I hope I have made a convincing case that they are, at least as long as one accepts that the data comes from something not too dissimilar from the forward model described here.

One type of misspecification that I don't expect the GLM to tolerate well is if the true gain control depended on something very different than the inverse of the contrast, as the $1/\sigma$ dependence is baked in in how we define the predictors. This doesn't seem like a very likely possibility, though; moreover, the general method described here could be easily extended if one had reason to believe that some other function of the contrast was important for the gain.

3 Examples

In this section, I give three examples of simulation and GLM fit. These are all done using the same parameters, given in Table 1. The only difference between the three simulations is the value of ξ, which controls the strength of the gain control.

These examples can be reproduced with the code I wrote by just running

```
% Instantiate a simulation with the default settings
s = Simulation();
% Run the simulation and generate plots
s.plot();
```

(see `Simulation`'s docstring for more details).

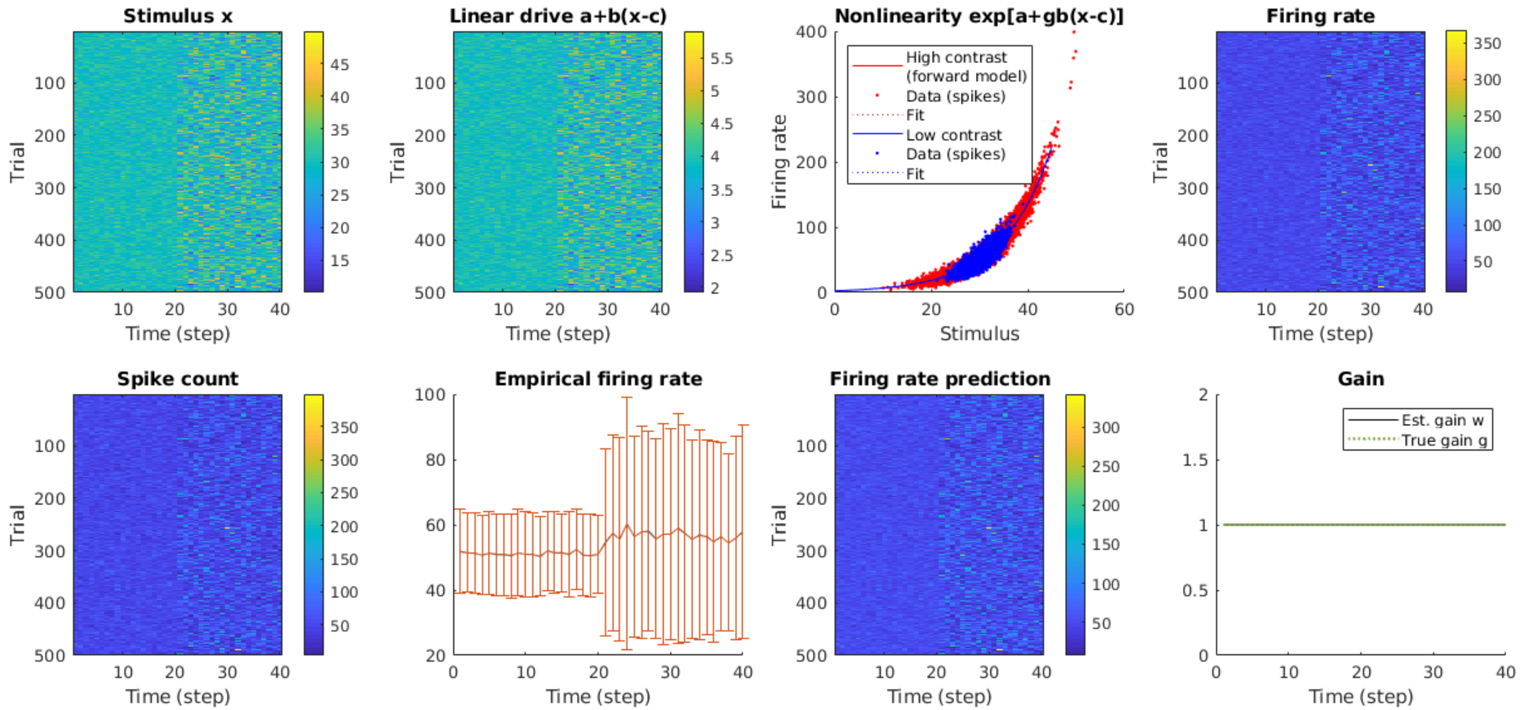

Figure 1: Low→high contrast simulation with $\xi = 0$ (no gain control). The second panel from the left in the bottom row gives the mean and standard deviation of the spike counts (mean and stddev taken across trials). Note how the variance of the firing rate increases after the contrast switch, and how both the true and estimated gain are fixed at $w = g = 1$.

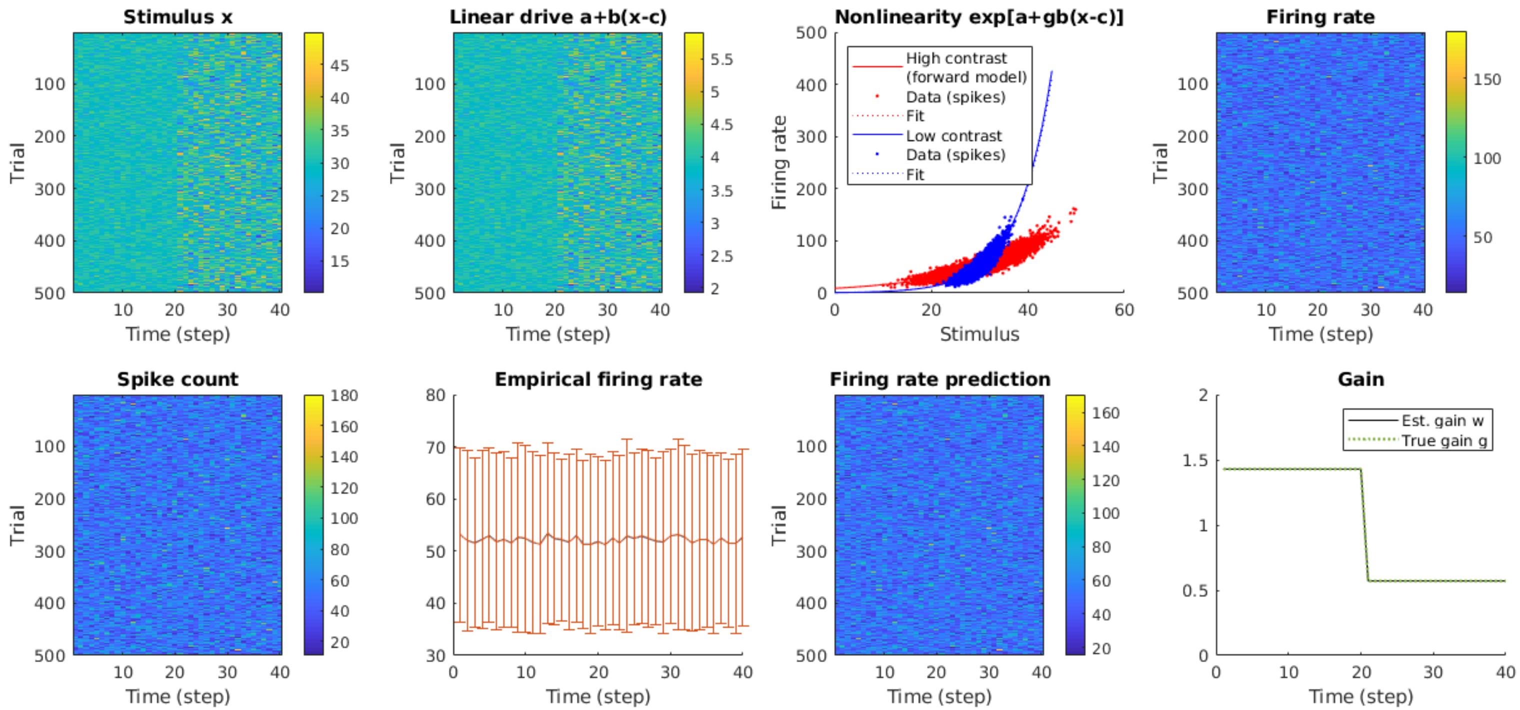

Figure 2: Low⟶high contrast simulation with $\xi = 1$ (optimal gain control). Note how the variance of the firing rate is constant throughout the simulation, optimizing the use of the dynamic range of the neuron. Note also how the true and estimated gain drop after the contrast switch.

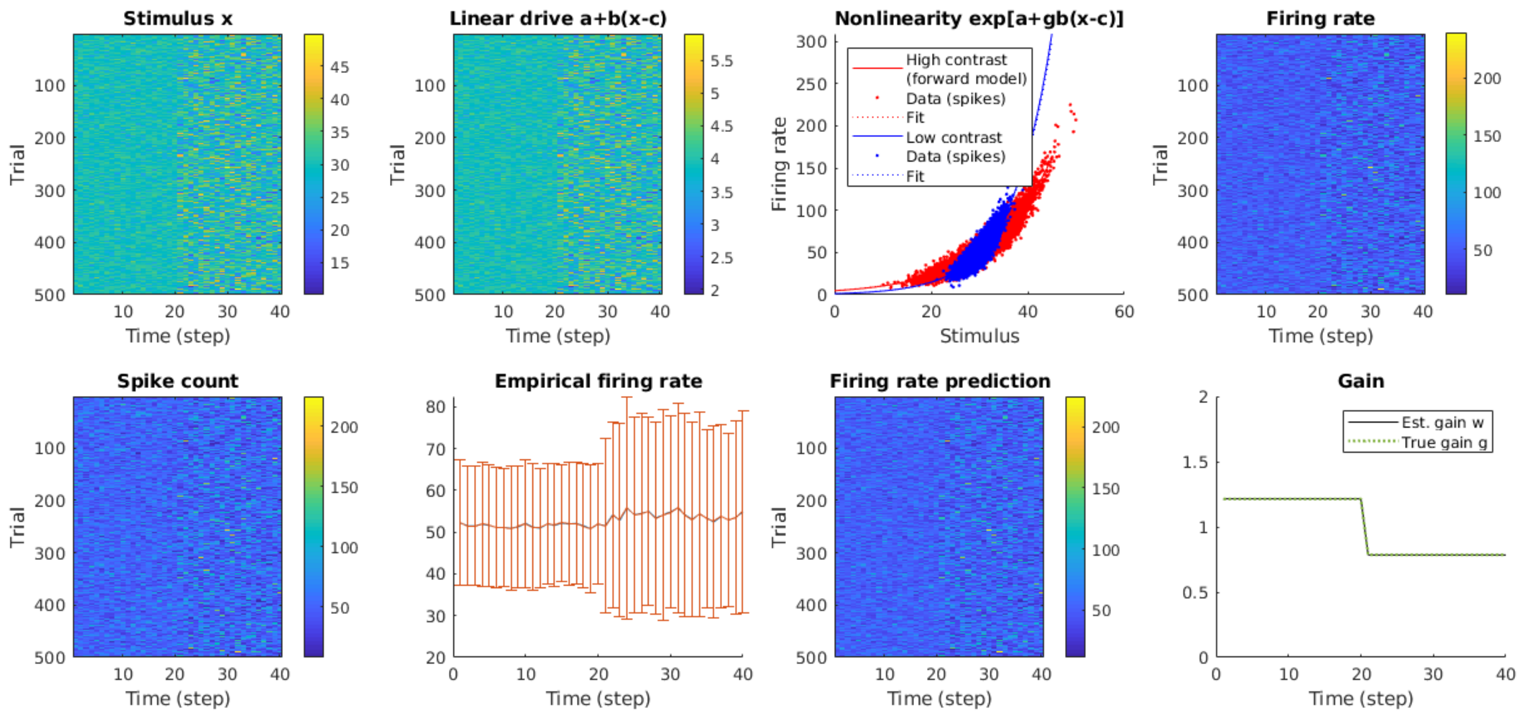

Figure 3: Low→high contrast simulation with $\xi = 0.5$. This is an intermediate case between those of Figure 1 and Figure 2.