

Recurrent generalized linear models with correlated Poisson observations

Marius Pachitariu, Lars Buesing, Maneesh Sahani

Summary

We introduce the Recurrent Generalized Linear Model (R-GLM), an extension of GLMs based on a compact representation of the spiking history through a linear recurrent neural network. R-GLMs match the predictive likelihood of Linear Dynamical Systems (LDS) with linear-Gaussian observations. We also address a disadvantage of GLMs, including the R-GLM, that they cannot model instantaneous correlations. The LDS however allows for extra correlated variability through the new innovation in the latent space. To improve GLMs we introduce a class of correlated output distributions which can be used with any type of multivariate data: binary, counts or continuous. The correlated Bernoulli distribution matches the predictive likelihood of Ising models for static binarized spike data. The correlated Poisson distribution offers significant improvements in predictive likelihood for GLMs and R-GLMs. We evaluate the performance of the models on a dataset recorded from a Utah array implanted into motor areas of a macaque monkey during a delayed reaching task. We report that the R-GLM consistently finds long timescales (of up to several seconds) of correlated activity similar to those found by LDS and longer than the timescales learnt by standard GLMs (up to 400 ms). Like all GLMs, the proposed model can be used with any link function and any output distribution. This is unlike models based on LDS which require careful approximations to be trained with Poisson outputs.

Additional Material

Over the past several years generalized linear models have been used extensively as statistical models of recorded neural data. However, it was recently shown that models based on latent dynamical systems can better capture the patterns of variability in spike trains [1]. This is primarily achieved through the long timescales of correlation which LDS models discover in neural data. These long timescales range from hundreds of milliseconds to seconds and are caused either by simultaneous external inputs which neurons receive or by the low-dimensional intrinsic dynamics of the neuronal networks in which they are embedded. To capture the long timescales with GLM-type models, we propose the R-GLM with the following parametrization:

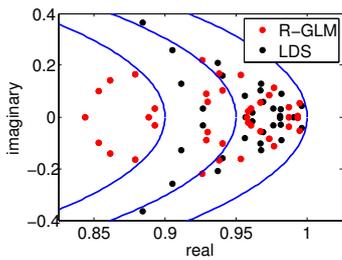
$$\mathbf{x}_1 = \mathbf{b}_1, \quad \mathbf{x}_{t+1} = A\mathbf{x}_t + W\mathbf{y}_t + \mathbf{b}_t, \quad \mathbf{y}_t \propto \text{Poisson}(f(C\mathbf{x}_t + Z\mathbf{y}_t + \mathbf{d})), \quad (1)$$

where \mathbf{y}_t is the vector of spike counts at time t , \mathbf{x}_t is a vector of hidden units with \mathbf{b}_1 its initial value and \mathbf{b}_t a driving input to the hidden units which provides a compact parametrization of the PSTH of the full population (see [1]). A , C , Z and W are various transformation matrices but Z is strictly lower triangular to get a well-defined process, \mathbf{d} is a vector of biases for each neuron and f a positive link function. The main difference from the LDS model is the lack of random innovations in the latent space of \mathbf{x}_t . Instead, \mathbf{x}_t is a deterministic function of the spiking history \mathbf{y}_t . The R-GLM effectively has no latent variables, so the probability of the observed spike trains can be written down explicitly in terms of the data and optimized by gradient descent. Although the resulting likelihood is not convex in R , training R-GLMs is very fast and stable in practice.

It was observed by [1] that an Ising-like model with Poisson outputs cannot be normalized if any of the interactions are positive, which greatly limits their usefulness in modelling instantaneous correlations. Also proved in [1], direct instantaneous couplings in a GLM result in ill-defined probabilistic models. Here we propose to model instantaneous correlations through the strictly lower triangular matrix Z in equation 1. The process \mathbf{y}_t given by equation 1 is well defined because the mean of y_{tj} only depends on the values y_{ti} up to $i = j - 1$. We call the distribution on \mathbf{y}_t correlated Poisson. If \mathbf{y}_t was continuous and the noise Gaussian, the model would be equivalent to full covariance Gaussians. In analogy we can define a correlated Bernoulli distribution. When trained on binarized data, the correlated Bernoulli distribution matches the performance of Ising models. The two models gave almost exactly the same likelihoods on test data for all cross-validation folds, with the correlated Bernoulli slightly but consistently better. Furthermore, the two models have the same number of parameters.

To compare with LDS, we changed the observation process of the R-GLM to Gaussian and set the link function to identity. We include a standard coupled GLM model also with Gaussian observations, where the spike counts of each neuron are projected onto a set of five basis functions before being linearly combined

across neurons [1]. This results in a total of $5q^2$ interaction parameters, with q the number of neurons, which we regularize by the L1 norm to obtain good generalization on held-out data. The basis functions we used were decaying exponentials with timescales .1, 20, 40, 100 and 400 ms (longer timescales did not help).



For LDS and R-GLM we followed [2] and L1-regularized the matrix A away from the identity matrix. This resulted in stable models and better predictive likelihoods. Furthermore, we initialized the LDS optimization with the new method based on subspace identification proposed in [3] which yielded significant improvements over common initialization techniques.

Figure 1 shows the eigenvalues of A for R-GLM and LDS trained on four minutes of data binned at 10ms. We included in the analysis 85 units. The longest timescales obtained by the two models were 2.6 and 3.4 seconds respectively. The average log likelihood improvement of both R-GLM and LDS with 40 hidden units over the standard GLM was 0.025 bits/spike. Using self-coupling filters in the R-GLM further improved it by 0.01 bits/spike. The dynamic models offer more compact parameterizations which helps their generalization performance. We next evaluated R-GLMs and standard GLMs with correlated Poisson outputs, driving inputs and self-coupling filters to model refractory effects. Table 1 shows in each line the improvement over the model in the line above. The proposed correlated Poisson distribution and the recurrent parametrization improve GLMs.

	Likelihood (bits/spike)
fully independent	- 3.15
correlated Poisson	+ 0.175
GLM	+ 0.225
GLM with correlated Poisson	+ 0.03
R-GLM with correlated Poisson	+ 0.03

To further show that the R-GLM better models the joint distribution of spike trains than GLMs, we compared the quality of the samples generated from the two models after training. The standard GLM was unstable at its best predictive likelihoods and the process diverged on 5-20% of simulated trials. On those trials where it did not diverge, the average correlation coefficients were too high and were similar to those shown in [1]. In contrast the R-GLM modeled well both instantaneous and time-lagged correlations between neurons. We followed the procedure of [1] and separated neurons into four non-overlapping groups based on

their average correlation coefficients to other neurons. Within each group we then averaged all pairwise correlations at different time lags. We computed these coefficients for data simulated from the trained R-GLM and found that they matched very well the time-lag correlations of the real data. Models based on LDS are also able to capture these statistics as shown in [1] on the same dataset. To conclude, R-GLMs bring to the GLM toolkit three essential properties enjoyed by latent dynamical systems: they have a compact parametrization that prevents overfitting, they can model dynamics and they can capture long timescales in a stable way. Furthermore, R-GLMs are quite modular and should be easy to extend along various dimensions.

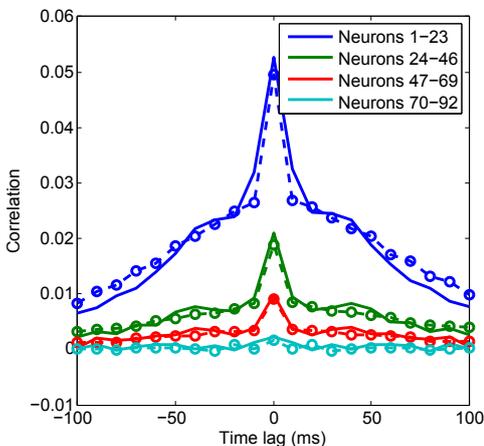


Figure 2: Average time-lag pairwise correlations within four disjoint groups of neurons. Continuous lines represent coefficients computed from real data.

Acknowledgements. Neural data from motor cortex were recorded in Krishna V. Shenoy’s laboratory.

- [1] JH Macke, L Busing, JP Cunningham, BM Yu, KV Shenoy, and M Sahani. Empirical models of spiking in neural populations. *Advances in Neural Information Processing Systems*, 24, 2011.
- [2] L Buesing, JH Macke, and M Sahani. Learning stable, regularised latent models of neural population dynamics. *Network: Computation in Neural Systems*, 23.
- [3] L Buesing, JH Macke, and M Sahani. Spectral learning of linear dynamics from generalised-linear observations with application to neural population data. *Advances in Neural Information Processing Systems*, 25, 2012.