

# Gaussian moment-closures for autoregressive GLM population models

M. Rule

January 8, 2019

In “[Autoregressive Point Processes as Latent State-Space Models](#)” we used moment-closures to coarse grain Autoregressive Point-Process Generalized Linear Models (AR-PPGLMs) of neural activity. These notes generalize these approximations for AR-PPGLM [models of neural ensembles](#). For networks of linear-nonlinear “neurons”, the derivations are very similar to the single-neuron case presented in the paper.

## 0.1 Motivation

Noise and correlations are important for modelling finite sized stochastic systems, such as neural networks. We usually model the noise in spiking neural models as conditionally Poisson. Sometimes, this noise is as important for capturing the dynamics as the mean firing rates.

Capturing noise and correlations is important in nonlinear systems, since noise can interact with nonlinearities to change the mean dynamics. Corrections for finite-size fluctuations and correlations have been explored theoretical in neural field models. The role of noise and correlations in the neural dynamics observed in large-scale neural population recordings has been less explored.

Here, we extend Rule and Sanguinetti (2018) to the case of populations. We outline an approximate second-order model based on a locally-quadratic Gaussian moment closure (i.e. cumulant expansion out to second order).

## 0.2 Autoregressive point-process models

Consider a stochastic process that counts the number of spikes that have occurred in neuron  $i$  up until time  $t$ :

$$N(i, t) = \text{spikes of neuron } i \text{ up to time } t \quad (1)$$

This process increments in discrete jumps. Its time derivative corresponds to a model of population spiking  $y(i, t)$  where each spike at time  $t_k$  is modelled as the derivative of the step function, (i.e. a Dirac delta distribution or impulse):

$$y(i, t) = \frac{d}{dt}N(i, t) = \sum_k \delta(t - t_k) \quad (2)$$

The ARPPGLM models the probability of increments in  $N$ , i.e. the probability of emitting spikes in  $y$ , as a conditionally-Poisson process based on a linear weighting of the process history, passed

through a rectifying pointwise nonlinearity  $\phi(\cdot)$ . This continuous-time conditional rate is called the conditional intensity,  $\lambda(i, t)$ .

In these notes, we conflate discrete and continuous-time notation as convenient. The math is identical and there is little practical difference between the two. We treat derivatives and integrals in the [Radon–Nikodym](#) sense, so  $\partial_t$  refers to either continuous or discrete-time differentiation, and  $\int_t$  referse to either discrete or continous time integration, as well as summation over neuronal indexes. For compactness, we denote functions that vary over time with subscripts, e.g.  $\lambda_{i,t}$ .

To define a population autoregressive GLM, define linear coupling weights  $\eta_{i,j,\tau}$  that capture how a spike in neuron  $j$  at time-lag  $\tau$  in the past influences neuron  $i$ . We also define a constant, mean input  $m_i$ , and an external time-varying input  $J_{i,t}$ , for each neuron. Overall, this population AR-PPGLM model can be defined as:

$$\begin{aligned}\lambda_{i,t} &= \phi \left( m_i + J_{i,t} + \iint_{\tau,j} \eta_{i,j,\tau} \cdot y_{j,t-\tau} \right) \\ y_{i,t} &\sim \text{Poisson}(\lambda_{i,t}),\end{aligned}\tag{3}$$

where  $\iint_{d\tau,dj} \dots$  denotes the integral over the population history, weighted by a linar history filter  $\eta_{i,j,\tau}$ .

### 0.3 Define an auxiliary population history process

As in Rule and Sanguinetti (2018), we convert the history dependence of the autoregressive point process into an auxiliary history process. This is a notational convenience that allows us to treat the whole system as Markovian (i.e. memoryless). This is similar to delay-line embedding in signal processing.

Define this history process as  $x_{\tau,i,t}$ . This history process contains a memory of spiking at time  $t - \tau$  at index  $\tau$ . The process evolves as a delay-line, moving backwards one time-step  $\tau$  per unit time  $t$ , and incorporating new spikes at time-lag zero (i.e.  $\delta_{\tau=0}$ ):

$$\begin{aligned}x_{\tau,i,t} &= y_{i,t-\tau} \\ \partial_t x_{\tau,i,t} &= -\partial_\tau x_{\tau,i,t} + \delta_{\tau=0} y_{i,t}\end{aligned}\tag{4}$$

To simplify notation, denote the linear history operator  $\$ \iint$

If the network is large, and the history filters smooth enough to average over many spikes, the law of large numbers applies. We can therefore neglect the binary nature of spiking, and instead consider averages over many spikes, which may be treated as effectively continuous. for a Poisson point process, this amount to making a diffusive approximation:

$$y \approx \lambda \cdot dt + \sqrt{\lambda} \cdot dW,$$

where  $dW$  is the derivative of the standard Wiener process. The diffusive (Langevin) approximation of the population AR-PPGLM is therefore:

$$\begin{aligned}
\lambda &= \phi(\mathbf{H}x + b) \\
y &= \lambda \cdot dt + \sqrt{\lambda} \cdot dW \\
\partial_t x &= -\partial_\tau x + \delta_{\tau=0} y
\end{aligned} \tag{5}$$

#### 0.4 The mean field case

In the mean-field case, we neglect the effects of noise entirely, and derive deterministic equations for the evolution of the mean rates  $\mu_y \approx \lambda$ :

$$\begin{aligned}
\mu_y &= \phi(\mathbf{H}x + b) \\
\partial_t \mu_x &= -\partial_\tau \mu_x + \delta_{\tau=0} \mu_y
\end{aligned} \tag{6}$$

This simplifies to

$$(\partial_t + \partial_\tau) \mu_x = \delta_{\tau=0} \phi(\mathbf{H}x + b). \tag{7}$$

#### 0.5 Moment-closure on the auxiliary history process

We can obtain closed (approximate) ordinary differential equations for the moments  $(\mu, \Sigma)$  of this process by considering a Gaussian moment closure of a quadratic approximation of this system about its current mean.

The evolution of the mean resembles the mean-field case, with an additional correction: There is a bias term, proportional to  $\frac{1}{2}\sigma^2$  times the curvature of the firing-rate nonlinearity,  $\phi''(\cdot)$ . (This is comparable to the effect of curvature in change of variables formula in [Itô's lemma](#).)

$$\begin{aligned}
\mu_y &= \phi(\mu_a) + \frac{1}{2} \phi''(\mu_a) \circ \sigma_a^2, \\
\mu_a &= \mathbf{H} \mu_x + b \\
\sigma_a^2 &= \text{Diag}(\mathbf{H} \Sigma_x \mathbf{H}^\top + \Sigma_b) \\
\partial_t \mu_x &= -\partial_\tau \mu_x + \delta_{\tau=0} \mu_y
\end{aligned} \tag{8}$$

Generally, one should consider the full Hessian of the nonlinearity,  $\nabla_x^2 \phi(\cdot)$ . However, since the nonlinearity is local, and doesn't entail any nonlinear interactions between neurons that aren't already accounted for by the linear terms, we only need to look at the diagonal variances of the activations  $a$ . If, perchance, one encounters a network where  $\phi$  includes pairwise interactions not captured in the linear terms, then the correction is:

$$\frac{1}{2} \text{tr}[\Sigma_a^\top \nabla_a^2 \phi(\mu_{a_i})]$$

For each neuron  $i$ , this is equivalent to taking the sum over the element-wise product of the the Hessian for  $\phi$  (for each neuron  $i$ ) and the covariance of the synaptic activations  $\Sigma_a = \mathbf{H} \Sigma_x \mathbf{H}^\top + \Sigma_b$ .

The covariance  $\Sigma_x$  evolves linearly according to the Jacobian  $\mathbf{J}$  of the mean system (Compare this to the [extended Kalman-Bucy filter](#)), and also experiences new noise from stochastic spiking,

denoted as  $\Sigma^{\text{noise}}(\mu)$ . This is one instance where the difference between continuous and discrete time matters. In continuous time

$$\begin{aligned}
\partial_t \Sigma_x &= \mathbf{J} \Sigma_x + \Sigma_x \mathbf{J}^\top + \Sigma^{\text{noise}}(\mu_x), \\
\mathbf{J} &= -\partial_\tau + \delta_{\tau=0} \nabla_x \mu_y \\
\nabla_x \mu_y &= \phi'(\mu_a) \cdot \mathbf{H} \\
\Sigma_{i,j,\tau_1,\tau_2}^{\text{noise}} &= \begin{cases} \mu_{y_{i,t}} & \text{if } i = j, \tau_1 = \tau_2 = 0 \\ 0 & \text{elsewise} \end{cases}
\end{aligned} \tag{9}$$

Since the noise is conditionally Poisson, the noise source term  $\Sigma_{\text{noise}}$  depends only on the mean  $\mu$ . As a corollary, every steady-state solution to the mean dynamics  $\mu_x$  is associated with exactly one steady-state solution to the correlations  $\Sigma_x$ . In this second-order Gaussian model, then, internally-generated correlations can change steady state values or stability, but don't lead to radically new dynamics. They mainly enter as a correction or modulation of the nonlinear, mean-field dynamics. This model can also capture how extrinsic noise affects the dynamics.

The second-order Gaussian moment closure is only valid as long as the quadratic approximation is accurate for a given covariance  $\Sigma$ . If the noise becomes large relative to the curvature of the nonlinearity  $\phi$ , this approximation fails. It is tempting to incorporate higher-order terms in the moment expansion in this scenario, but this is dangerous. Firing rate nonlinearities in neural models typically saturate, and become flat above or below a certain value. Polynomial approximations based on Taylor series expansions diverge for such functions. Instead, one should use a moment closure that takes into account how noise interacts with nonlinearity globally, such as the [Hennequin-Lengyel](#) moment closure for rectified quadratic nonlinearities.