Machine Learning Methods for Neural Data Analysis

Lecture 5: Deconvolution with a Point Process Prior

Scott Linderman

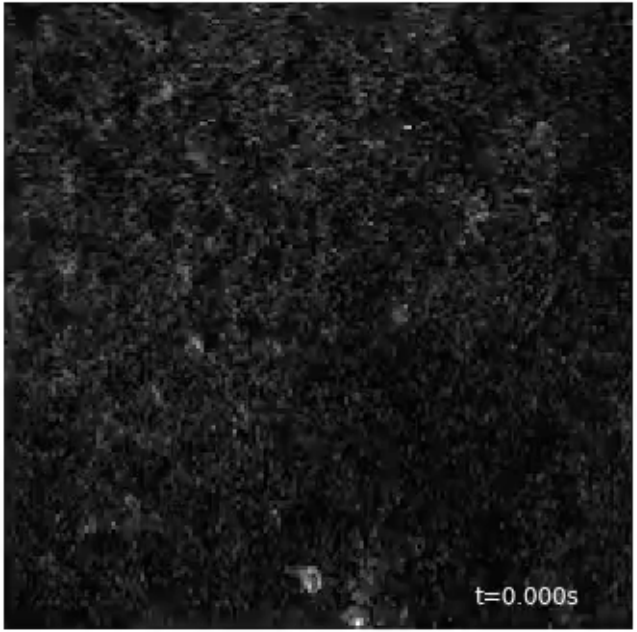
STATS 220/320 (NBIO220, CS339N). Winter 2021.

Announcements

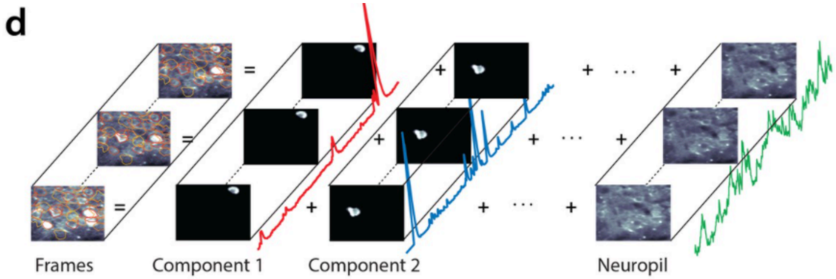
- Lab 2 due Thursday, 11:59pm.
- My office hours today 1-2pm.

Recap

2 photon calcium imaging



Sue Ann Koay and David Tank



$$Y = U^T C + u_0 c_0^\top + \epsilon$$

$$U \in \mathbb{R}^{N \times P} \; C \in \mathbb{R}_+^{N \times T} \; u_0 \in \mathbb{R}^P \; c_0 \in \mathbb{R}^T$$

$$\epsilon_{pt} \sim \mathcal{N}(0, \sigma^2)$$

Recap

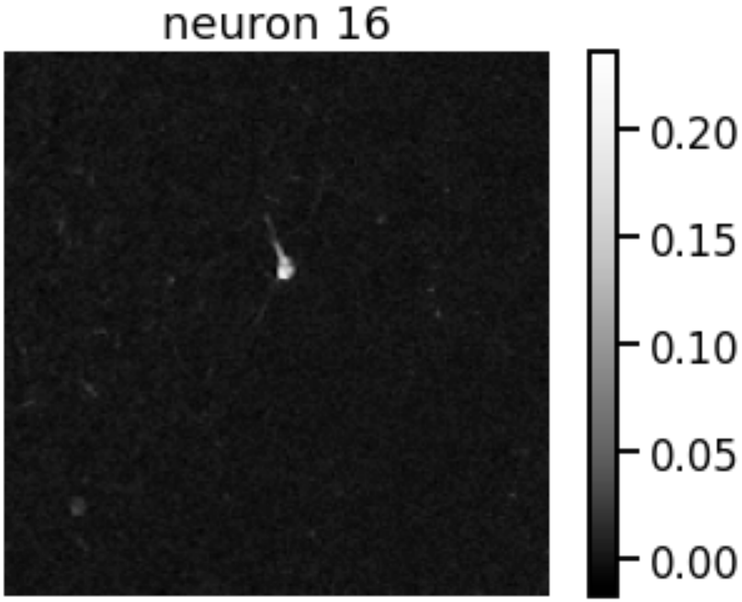
Optimizing the footprints

Given an estimate of the fluorescence trace c_n, the optimal footprint is,

$$u_n = \frac{R_n c_n}{\|R_n c_n\|}$$

where $R_n \in \mathbb{R}^{P \times T}$ is the residual.

This is assuming a uniform prior on unit-norm footprints.



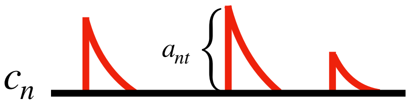
Recap

Prior on calcium traces

We had an exponential prior on spike amplitudes, i.e. jumps in the fluorescence:

$$p(c_n) = \prod_{t=1}^{T} \mathrm{Exp}(a_{nt})$$

$$= \prod_{t=1}^{T} \mathrm{Exp}(c_{nt} - e^{-1/\tau} c_{n,t-1}; \lambda_n)$$



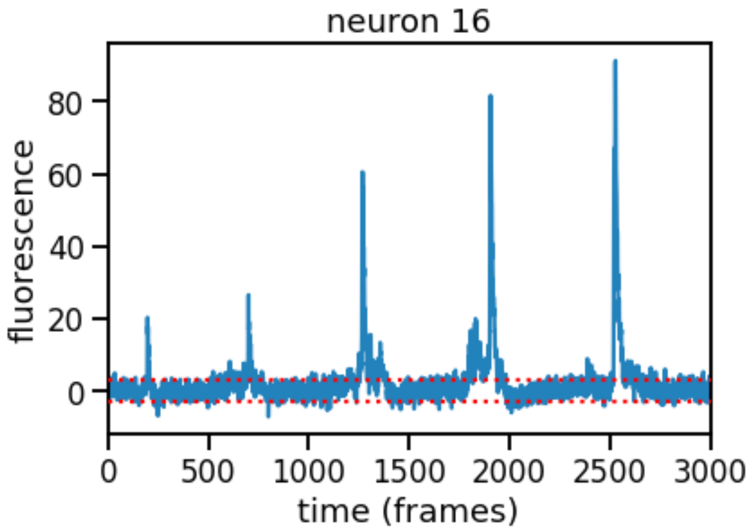
Recap

Objective for optimizing traces

Combining the log likelihood and the log prior lead to the following objective:

$$
\begin{aligned}
\mathscr{L}(c_n) &= -\frac{1}{2\sigma^2}\|c_n - \mu_n\|_2^2 + \lambda_n \sum_{t=1}^{T} (c_{nt} - e^{-1/\gamma} c_{n,t-1}) \\
&= -\frac{1}{2\sigma^2}\|c_n - \mu_n\|_2^2 + \lambda_n \|G c_n\|_1.
\end{aligned}
$$

where $\mu_n = R_n^\top u_n$ is the residual projected onto the spatial factor for this neuron and $Gc_n = a_n$ are the spike amplitudes.



Recap

Objective for optimizing traces

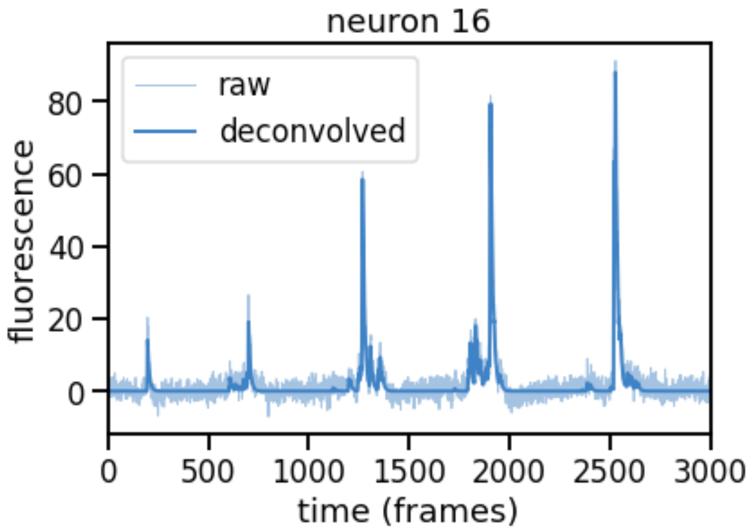
- Maximizing $\mathscr{L}(c_n)$ is equivalent to solving the following convex optimization problem,

$$\hat{c}_n = \text{arg min}_{c_n} \; \|Gc_n\|_1$$
$$\text{s.t. } \|c_n - \mu_n\|_2 \leq \theta$$
$$c_n \geq 0,$$

  for some threshold θ.

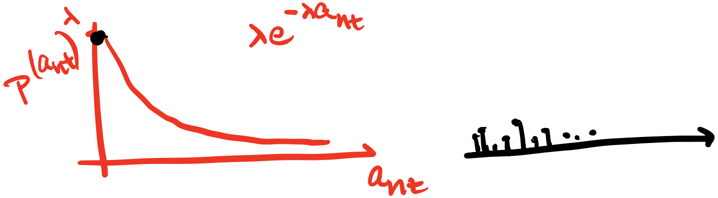
- **Idea:** set $\theta = (1 + \epsilon)\sigma\sqrt{T}$.

- Since G is a sparse banded matrix, we can solve this in $O(T)$ time with CVXpy. (Lab 3)

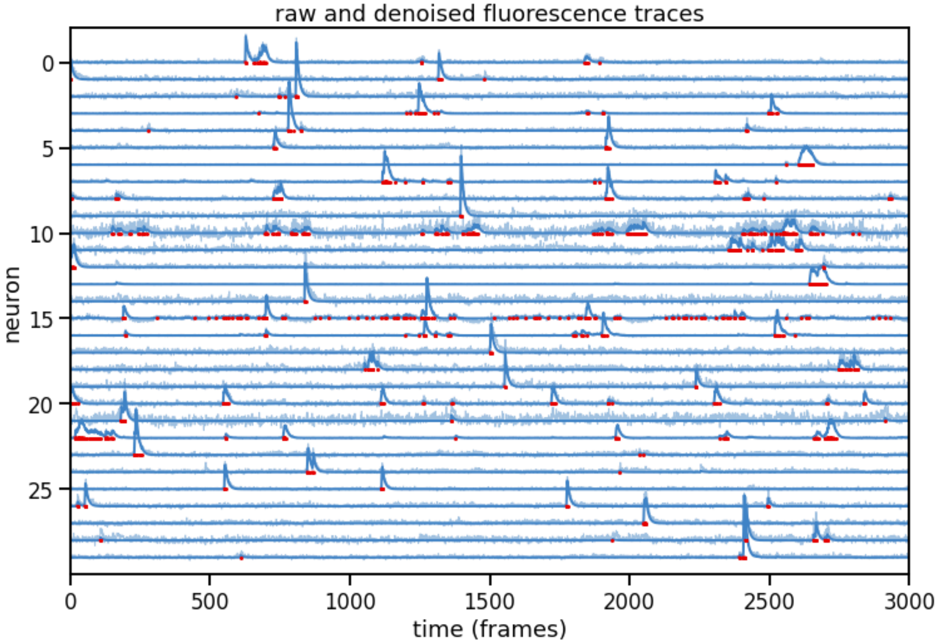


Recap

Limitations of “L1” deconvolution

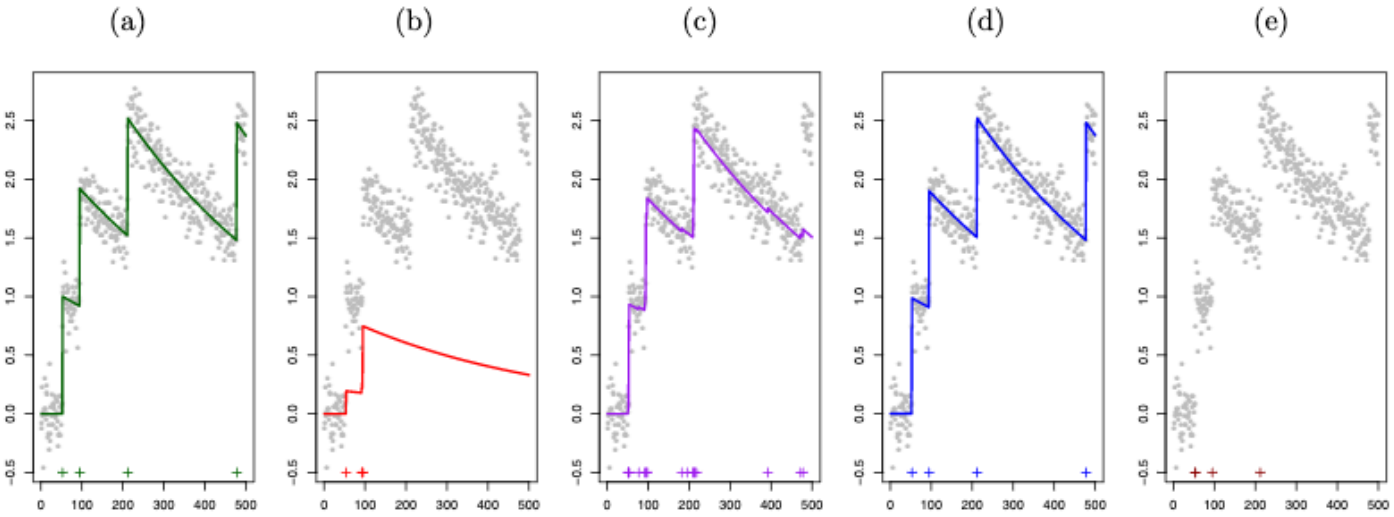


- The resulting traces are sparse but still overestimate the number of spikes.
- Likewise, the amplitudes are penalized, shrinking our estimates.
- Many of these issues can be traced back to the exponential prior distribution on amplitudes.



Recap

Limitations of “L1” deconvolution



Jewell and Witten, AOAS 2018

Agenda

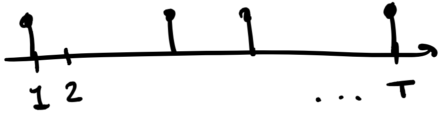
1. A point process prior
2. Efficient optimization via dynamic programming

Point processes

(More to come next week)

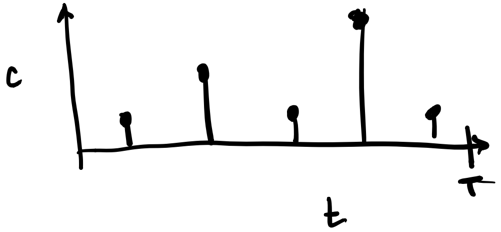
- distribution on sets $\mathcal{X} \subset \{1, \dots, T\}$ eg $\mathcal{X} = \{t_k\}_{k=1}^{K}$

$1 \le t_1 < t_2 < \dots < t_K \le T$



- $\mathcal{X} \subset \{1, \dots, T\} \times \mathbb{R}_+$

eg $\mathcal{X} = \{(t_k, c_k)\}_{k=1}^{K}$



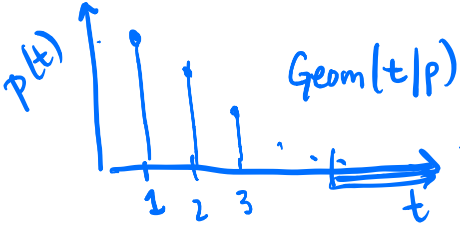
A point process prior for spikes

$$\mathrm{Unif}(c_{t_k} \mid [0, c_{max}])$$

$$X = \{(t_k, c_{t_k})\}_{k=1}^{K}$$

$$p(X) = \left[p(t_1) \prod_{k=2}^{K} [p(t_k \mid t_{k-1})] \times p(t_{K+1} > T) \right] \times \left[\prod_{k=1}^{K} p(c_{t_k}) \right]$$

$$\mathrm{Geom}(t_1 \mid p) \qquad \mathrm{Geom}(t_k - t_{k-1} \mid p) \qquad \sum_{t_{K+1}=T+1}^{\infty} \mathrm{Geom}(t_{K+1} - t_K \mid p)$$



$$\mathrm{Geom}(t \mid p) = (1-p)^{t-1} p$$

$$\Pr(t \le T \mid p) = 1 - (1-p)^{T}$$

$$p(X) = (1-p)^{t_1 - 1} p \prod_k (1-p)^{t_k - t_{k-1} - 1} \cdot p \, (1-p)^{T - t_K} \cdot \left(\frac{1}{c_{max}}\right)^K$$

$$= (1-p)^{T} \left(\frac{p}{1-p}\right)^{K} \cdot \left(\frac{1}{c_{max}}\right)^{K}$$

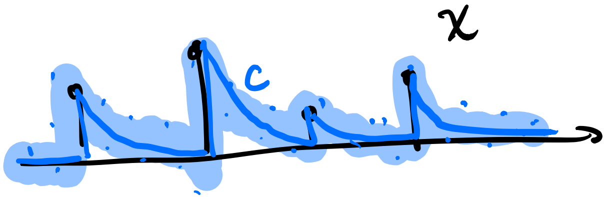
From points to calcium traces

Goodbye convolutional model!

$$p(x) \propto \left(\frac{\rho}{(1-\rho)\, c_{max}}\right)^{K}$$

$$x \rightarrow (c_1 \cdots c_T)$$

$$c_t(x) = \begin{cases} c_{t_k} & \text{if } t \in \{t_k\} \\ e^{-1/\tau}\, c_{t-1}(x) & \text{o/w} \end{cases}$$



Deriving the prior probability

Rewriting the likelihood

In terms of spike times and amplitudes

$$\log p(\mu \mid \chi) = -\frac{1}{2\sigma^2} \sum_{t=1}^{T} \left(\mu_t - c_t(\chi)\right)^2$$

$$= -\frac{1}{2\sigma^2} \sum_{k=0}^{K} \sum_{t=t_k}^{t_{k+1}-1} \left(\mu_t - c_{t_k} e^{-(t-t_k)/\tau}\right)^2$$

$$\left[t_0 = 1 \, , \; t_{K+1} = T+1 \right]$$

Defining our objective

$$\mathcal{L}(x) = \log p(x) + \log p(\mu \mid x)$$

$$\log \frac{p}{(1-p)\,c_{max}}$$

$$= \frac{-1}{2\sigma^2} \sum_{k=0}^{K} \sum_{t=t_k}^{t_{k+1}-1} \left(\mu_t - c_{t_k} e^{-(t-t_k)/\tau} \right)^2 + \eta K + \text{const}$$

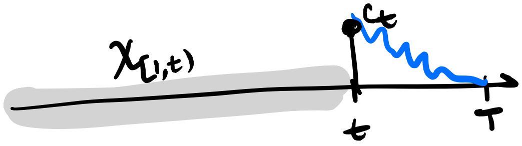
GOAL: $x^* = \operatorname{argmax} \mathcal{L}(x)$

Optimizing the spikes with dynamic programming

This looks hard…

- $\mathscr{X}$ is a set of unknown cardinality
- Each entry consists of a time and an amplitude
- **Idea:** pull the last spike out of $\mathscr{X}$

Rewrite the objective in terms of $\mathcal{X}_{[1,t)}$, t, and c_t



$$\mathcal{L}(\mathcal{X}) = \mathcal{L}\left(\mathcal{X}_{[1,t)} \cup \{(t, c_t)\}\right)$$

$$\hat{=} \mathcal{L}\left(\mathcal{X}_{[1,t)}, t, c_t\right)$$

$$= \mathrm{LOG}\; p\left(\mu_{[1,t)}, \mathcal{X}_{[1,t)}\right) - \frac{1}{2\sigma^2} \sum_{t'=t}^{T} \left(\mu_{t'} - c_t e^{-(t'-t)/\tau}\right)^2 + \eta$$

Optimizing wrt the amplitude c_t

Optimizing wrt previous spikes, $\mathscr{X}_{[1,t)}$

It's the same problem on a subset of data!

Optimizing wrt the time of the last spike, t

Putting it all together

Considerations

- Complexity?
- Hyperparameters?

Conclusion

- **Point processes** offer a truly sparse prior.
- While they seem to pose a hard **combinatorial optimization problem**, in fact we can solve for the optimal spikes in polynomial time.
- This is what the Allen Institute currently uses in their 2P analysis pipelines.

Further reading

- Ch 3 of the course notes (coming later today)
- Jewell, Sean, and Daniela Witten. 2018. "Exact Spike Train Inference via ℓ_0 Optimization." The Annals of Applied Statistics 12 (4): 2457–82.