A reproduction and review of *Exact spike train inference via ℓ_0 optimization*

Yiqun Chen*

*yiqunc@uw.edu

Abstract

Advances in neuroscience in the past decade, including multi-electrode arrays and Calcium imaging, have enabled scientists to collect single-neuron resolution data of an unparalleled magnitude and precision. The bottlencecks in neuronscience discovery, therefore, no longer lie simply in collecting data from neural populations, but also in processing and analyzing such data.

Calcium imaging technique, in particular, is able to measure fluorescence levels of individual cells simultaneously, where the measured traces can be seen as a first-order approximation of cells' activities over time. Determining the exact time of the "spikes" for individual neurons is of great scientific interest and remains an open problem in the field of neuroscience.

A recent and popular attempt in [Friedrich et al., 2017] formulated the problem into a convex optimization problem with an ℓ_1 penalty. In this paper we consider a modification with an ℓ_0 penalty which was previously thought to be computationally intractable. In contrast to the conventional wisdom, we use a dynamic programming (DP) algorithm to achieve the global optimum efficiently. In addition, simulations and real data analyses demonstrated superiority in spike detection performance.

While we dedicate the majority of our discourse to reproducing the paper, we also include more proof details and a few more simulation settings omitted by the authors. Unoptimized python code and Jupyter notebook for generating the results of our paper can be found in the listed `GitHub` repository: https://github.com/yiqunchen/biostat_572_paper.

Introduction

Understanding neuronal dynamics during information processing is central for the understanding of higher brain functions [Tian et al., 2009, Stosiek et al., 2003, Glimcher, 2011, Reid, 2012, Paninski and Cunningham, 2018]. However this question was deemed as impossible to answer not only because of the immense number of neurons in the brain, but also due to the lacking technology in monitoring real-time neuronal activities *in vivo*. Much to the field's surprise, the past decade has seen the rapid development of Calcium imaging, which has become one of the most used techniques for recording activities from large neural populations *in vivo* [Tian et al., 2009, Chen et al., 2013, Dombeck et al., 2007, Zeng and Sanes, 2017, Pachitariu et al., 2018, Paninski and Cunningham, 2018]. The basic principle of Calcium imaging is that neural activities (often referred to as "spikes") cause rapid changes in intracellular free Calcium and one can rely on this principle to probe the excitations of imaged neurons by co-expressing the neuronal cells with genetically encoded Calcium indicators or protein sensors. Such sensors can be targeted to specific cell types for non-invasive imaging of identified neurons over long timescale, minutes or even hours at a

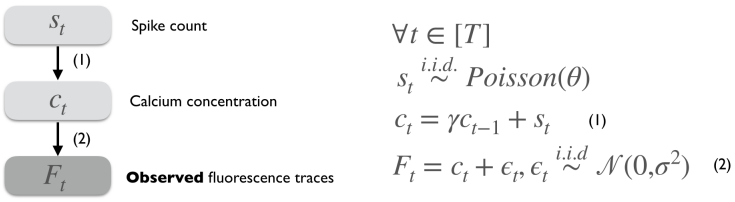

Figure 1. Data generating process for Calcium imaging data where we use light grey to denote unobserved variables and dark grey for observed variables

30-60 Hz sampling frequency. [Tian et al., 2009, Chen et al., 2013, Smetters et al., 1999, Carsen et al., 2018, Pachitariu et al., 2018].

However, unlike for electrophysiological data, standardized algorithms and analysis pipelines for Calcium imaging data have not converged (or even emerged, as some would argue). One of the many challenges of extracting the underlying spike process from the fluorescence signal is that single image pixel can contain a complex mixture of signals from neurons, neuropil, and noise, which then are temporally convolved to produce an intracelluar calcium dynamics [Carsen et al., 2018, Pachitariu et al., 2018, Stringer and Pachitariu, 2018, Mukamel et al., 2009]. Disentangling the signal from noise in a principled fashion is non-trivial and many have attempted to address the problem with different approaches including template matching, reduction to supervised learning, sparse non-negative deconvolution, etc [Pnevmatikakis, 2018, Paninski et al., 2007, Paninski and Cunningham, 2018, Paninski and Cunningham, 2018]. A recent algorithm introduced in [Friedrich et al., 2017] proposed an online active set method to infer spikes (OASIS) for an order-1 autoregressive (AR-1) data generating process with ℓ_1 penalty.

In this paper, therefore, we consider an AR-1 process similar to those considered by many in the field [Friedrich et al., 2017, Paninski and Cunningham, 2018, Vogelstein et al., 2010]. A graphical representation of the data generating mechanism can be found in Figure 1. In other words, we posit that at each time point neuronal spiking follows a Poisson process s_t which changes the intracellular Calcium concentration c_t. The latter, in the absence of s_t, follows an autoregressive-1 process; finally, we observe an noisy realization F_t of the concentration c_t.

Note that we only observe F_t and the goal is to estimate s_t and we note that in general the fluorescence can depend on calcium concentration in an affine manner, i.e., $F_t = \beta_0 + \beta_1 c_t$ for some β_0, β_1. However, for all our intensive purposes, we will proceed with assuming that $\beta_0 = 0$ and $\beta_1 = 1$. The former two assumptions are in principle made without loss of generality since one can estimate β_i's from data or use a grid of plausible β_i values and incorporate them into the estimation pipeline outlined in the method section [Jewell and Witten, 2018, Jewell et al., 2019].

Two of the most popular spike inference algorithms ([Friedrich et al., 2017] and [Vogelstein et al., 2010]) considered a generative model where s_t are interpreted as the number of spikes, which naturally leads to an MAP (maximum a posteriori) estimator of s_t. While in principle a count valued distribution such as the Poisson distribution as a prior on s_t would be most consistent, the MAP estimator $\hat{s}_t$ in that case would lead to a computationally intractable problem (See Appendix for more details). Therefore some have used an exponential prior on s_t instead and arrive at the non-negative factorization problem in [Vogelstein et al., 2010, Friedrich et al., 2017] and we will henceforth refer to this

optimization problem as the ℓ_1 problem.

$$\begin{aligned} \min_{c_1,\cdots,c_T,s_1,\cdots,s_T\in\mathbf{R}} \quad & \sum_{t=1}^{T}(F_t - c_t)^2 + \lambda|s_t| \\ \text{subject to} \quad & s_t = c_t - \gamma c_{t-1}, \forall t \in [T] \\ & c_t - \gamma c_{t-1} \geq 0, \forall t \in [T] \end{aligned} \tag{1}$$

A variant of the ℓ_1 problem will penalize the initial concentration $|c_1|$ as well; since in practice this additional penalty does not make a substantial difference, we will, by abuse of our terminology, treat these two variants as essentially the same problem. Note that both variants are convex problems and therefore we have a set of efficient algorithms which obtain the global minimum of the objective [Friedrich et al., 2017, Boyd and Vandenberghe, 2004, Hastie et al., 2015]. Since the output $\hat{s}_i$ will not be integer valued anymore, one cannot interpret it as the number of spikes at each time point; instead, a larger value of $\hat{s}_i$ is often interpreted as a higher confidence that one or more spikes have occurred at a specific time point. In reproducing the chosen paper, we consider a closely related optimization problem that can be motivated in two different ways. First, we consider the ℓ_0 analog of the aforementioned optimization problem where $|s_t| = |c_t - \gamma c_{t-1}|$ is replaced with $1(s_t \neq 0)$. The names essentially come from the fact that these are ℓ_1 and ℓ_0 norms of the vector $s = (s_2, \cdots, s_T)$. Second, one can consider the same MAP problem but instead of putting a Poisson prior to it, we consider a Poisson prior with $s_t \in 0, 1$, which can be interpreted as whether textitat least one spike has occurred (See Appendix for more details). However the optimization problem one arrives at is highly non-convex and computationally intractable. However, the contribution of this paper is that by dropping the last non-negative constraint $s_t = c_t - \gamma c_{t-1} \geq 0$, we can solve the ℓ_0 optimization problem using a computationally efficient dynamic programming (DP) algorithm. We remark that in practice the non-negative constraint makes little difference unless recorded fluorescence traces are extremely noisy.

$$\begin{aligned} \min_{c_1,\cdots,c_T,s_1,\cdots,s_T\in\mathbf{R}} \quad & \sum_{t=1}^{T}(F_t - c_t)^2 + \lambda 1(s_t \neq 0) \\ \text{subject to} \quad & s_t = c_t - \gamma c_{t-1}, \forall t \in [T] \\ & c_t - \gamma c_{t-1} \geq 0, \forall t \in [T] \end{aligned} \tag{2}$$

To further motivate the ℓ_0 problem, we consider a proof of concept analysis where we compare and contrast the behavior of ℓ_0 and ℓ_1 solutions under the data generating mechanisms outlines in Figure 1. We can see from Figure 2 that ℓ_0 solutions approximate both the true spiking times and calcium concentration well, which suggest that calcium concentration fitting could potentially be used to select an optimal penalty λ. Regarding the ℓ_1 solutions, however, we observe a mismatch in good curve fitting and spike estimation: we either tend to overestimate the number of spikes or underfit the observed fluorescence. We note that as remarked in [Friedrich et al., 2017], post-thresholding on the estimated $\hat{s}_t$ will increase the accuracy of spike estimation, which are shown in orange bars in the rightmost panel of Figure 2. This however involves additional freedom parameter which can be hard to choose without access to the locations of true spikes.

The rest of our paper will be organized as the following: we will first introduce our modified optimization problem, present a DP algorithm, and prove its correctness. Then we will present the results comparing ℓ_1 and ℓ_0 algorithms on simulated datasets. Finally we conclude with an application section on real recordings, which demonstrates excellent result relative to both ground truth and ℓ_1 estimates. Throughout the paper we will remark on places where our results and approaches differ from those of the original paper.

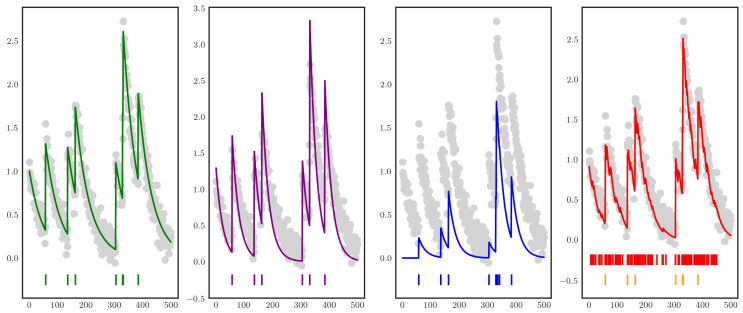

Figure 2. Comparing ℓ_0 and ℓ_1 results with different penalizations under the data generating mechanisms. The first panel (green) corresponds to the true Calcium concentration and the spike times; there are 6 spikes in the 500 time points. The second plot (purple) is generated by solving the ℓ_0 problem with a tuning parameter (λ) which estimates 6 spikes. Similarly, the third plot (blue) is generated by solving the ℓ_1 problem with λ which estimates 7 spikes and the fourth plot (red) is generated by using λ which generates the best approximation in Calcium concentrations.

Methods

Proposed algorithm

We first recall the optimization problem we aim to solve below.

$$G(T;\lambda) = \min_{c_1,\cdots,c_T \in \mathbf{R}} \quad \sum_{t=1}^{T} (F_t - c_t)^2 + \lambda 1(c_t - \gamma c_{t-1} \neq 0) \tag{3}$$

Note that we have dropped the non-negative constraint, which as we will see in a moment, permits us to devise an efficient algorithm which attains the global minimum of the objective $G(T)$. We have also eliminated the equality constraints and the variable s_t by substitution; finally $\lambda > 0$ is a non-negative tuning parameter penalizing the number of estimated spikes.

To build up to our DP algorithm, we first make the following two observations about this problem:

- Given that two spikes are present at time $t_1 < t_2$, the Calcium concentration decays exponeitally with decaying coefficient γ and the estimation can be performed in $O(t_2 - t_1)$ steps using a least square algorithm.
- Given that a spike has occurred at time point t, the estimated Calcium concentration at any time $t_1 < t$ will not be a function of the estimated Calcium concentration for $t_2 > t$; this follows easily from the fact that condition on $c_t - \gamma c_{t-1} \neq 0$, the optimization problem decouples over the two segments (before and after t).

Given the aforementioned observations, it suffices to estimate the time point t such that $c_t - \gamma c_{t-1} \neq 0$ and we remark that this is equivalent to a rich set of well-studied problems referred to as the changepoint detection problem. Among the extensive work detected to changepoint detection, [Auger and Lawrence, 1989] and [Jackson et al., 2003] considered a DP algorithm which solves the general version of the ℓ_0 problem and we outlined the details of the algorithm in Algorithm 1.

Algorithm 1 DP Algorithm for the ℓ_0 problem

Input : Input fluorescence sequence $F_t, t \in [T]$, Parameters $\gamma \in (0,1)$, $\lambda > 0$
Output : Estimated calcium concentration $\hat{c}_t, t \in [T]$, estimated spikes $\hat{s}_t$
Initialize $cp(0) = \emptyset$, $G(0) = -\lambda$
for $t = 1, \cdots, T$ **do**
  Compute $G(t) = \min_{0 \le s < t} \{G(s) + \mathcal{D}(F_{s:t}; \gamma) + \lambda\}$
  Set $t' = \arg\min_{0 \le s < t} \{G(s) + \mathcal{D}(F_{s:t}; \gamma) + \lambda\}$
  Update $cp(t) = cp(t') \bigcup t'$
end
$\hat{s}_t = 1(t \in cp(T))$

$$\hat{c}_t = \begin{cases} \gamma \cdot \hat{c}_{t-1} & \hat{s}_t = 0 \\ \arg\min_y \sum_{t'=t}^{\min \tilde{t} > t : \hat{s}_{\tilde{t}} \neq 0} (F_{t'} - \gamma^{t'-t} \cdot y)^2 & \hat{s}_t \neq 0 \end{cases}$$

Before we prove the correctness of the dynamic programming algorithm, we introduce the following notations for a less cumbersome presentation:

- $cp(t), t \in [T]$ will be the set of change points up to time step t (exclusive).
- Given a fluorescence trace $F_{t_1:t_2}$ and decay parameter γ, we define

$$\mathcal{D}(F_{t_1:t_2}; \lambda) = \min_{c_t \text{ s.t. } c_{t+1} = \gamma c_t} \frac{1}{2} \sum_{t=t_1}^{t_2} (F_t - c_t)^2 \tag{4}$$

  Upon inspection, to compute $\mathcal{D}(F_{t_1:t_2}; \lambda)$, it amounts to solve a least square problem which we claim to have $O(t_2 - t_1)$ time complexity, formalized below.

Lemma 1 (Computation of $\mathcal{D}(F_{t_1:t_2}; \lambda)$)**.** *Given F and λ, $\mathcal{D}(F_{t_1:t_2}; \lambda)$ has the following closed form solution:*

$$\mathcal{D}(F_{t_1:t_2}; \lambda) = \frac{1}{2} \sum_{t=t_1}^{t_2} (F_t - \gamma^{t-t_1} \cdot \hat{c}_0)$$

where

$$\hat{c}_0 = \frac{\sum_{t=t_1}^{t_2} \gamma^{t-t_1} \cdot F_t}{\sum_{t=t_1}^{t_2} \gamma^{2(t-a)}}$$

Proof. First, by eliminating the equality constraint, one can rewrite the optimization problem as

$$\mathcal{D}(F_{t_1:t_2}; \lambda) = \min_{c_0} \frac{1}{2} \sum_{t=t_1}^{t_2} (F_t - \gamma^{t-t_1} c_0)^2 \tag{5}$$

.

Now this is simply a least square problem so taking derivative with respect to c_0 and setting the derivative to 0 gives us the optimal solution.

Firs order condition: $\sum_{t=t_1}^{t_2} (F_t - \gamma^{t-t_1} c_0) \cdot (\gamma^{t-a}) = 0 \implies \hat{c}_0 = \frac{\sum_{t=t_1}^{t_2} \gamma^{t-t_1} \cdot F_t}{\sum_{t=t_1}^{t_2} \gamma^{2(t-a)}}$.

□

Theorem 2 (Correctness of Algorithm 1)**.** *Algorithm 1 solves the ℓ_0 optimization problem in $O(T^2)$ operations*

Proof. We first prove the correctness of our algorithm.

$$G(T) = \min_{c_1, \cdots, c_T} \frac{1}{2} \sum_{t=1}^{T} (F_t - c_t)^2 + \lambda \sum_{t=2}^{T} 1\{c_t \neq \gamma c_{t-1}\}$$

$$\overset{1.}{=} \min_{|\mathcal{P}|=K\in\mathbb{N}\,\forall t\in[T]\setminus\mathcal{P}, c_t=\gamma c_{t-1}} \min \frac{1}{2}\sum_{t=1}^{T}(F_t - c_t)^2 + \lambda\sum_{t=2}^{T} 1\{c_t \neq \gamma c_{t-1}\}$$

$$\overset{2.}{=} \min_{0<\tau_0<\cdots<\tau_K<\tau_{K+1}=T, K\in\mathbb{N}} \sum_{j=0}^{K}\Big\{\frac{1}{2}\min_{c_{\tau_j+1}}\Big[\sum_{t=\tau_j+1}^{\tau_{j+1}}(y_t - \gamma^{t-(\tau_j+1)}\cdot c_{\tau_j+1})^2\Big]\Big\} + \lambda\cdot K$$

$$\overset{3.}{=} \min_{0<\tau_0<\cdots<\tau_K<\tau_{K+1}=T, K\in\mathbb{N}} \sum_{j=0}^{K}\Big\{\mathcal{D}(y_{\tau_j} : y_{\tau_{j+1}}) + \lambda\cdot K\Big\}$$

$$\overset{4.}{=} \min_{s<T}\Big\{\min_{0<\tau_0<\cdots<\tau_K=s<T, K-1\in\mathbb{N}}\Big[\sum_{j=0}^{K-1}\mathcal{D}(y_{\tau_j} : y_{\tau_{j+1}}) + \lambda\cdot(K-1)\Big] + D(y_{\tau_K} : T) + \lambda\Big\}$$

$$= \min_{s<T}\Big\{G(s) + \mathcal{D}(s : T) + \lambda\Big\}$$

.

Now step 1. follows from rewriting the problem into a joint minimization over the number of places where $c_t \neq \gamma c_{t-1}$ and the estimation of c_i's, where the former was denoted by $\mathcal{P}$

Step 2. just follows from explicitly writing out K change points, i.e., for all τ_j, we have $c_{\tau_j+1} \neq \gamma c_{\tau_j}$.

3. follows from the definition of $\mathcal{D}(\cdot)$ where we suppressed the dependence on λ and 4. follows from expanding the sum into two parts.

Now note that we basically get the following recursive relation and can be unrolled using a simple for loop: $G(T) = \min_{s<T}\{G(s) + \mathcal{D}(s : T) + \lambda\}$.

In particular, we have the following steps:

- $G(0) = -\lambda$
- $G(1) = G(0) + \mathcal{D}(y_1) + \lambda$
- $G(2) = \min\{G(0) + \mathcal{D}(y_{1:2}) + \lambda, G(1) + \mathcal{D}(y_2) + \lambda\}$
- $\cdots$

Now the time complexity is simply $\sum_{t=1}^{T} t = O(T^2)$. □

A recent paper by [Killick et al., 2012] provided more efficient solutions to the changepoint detection with some assumptions on $\mathcal{D}(\cdot)$. In particular, the main insight can be summarized as the following: suppose that at a time point we have $s < r$ and $G(s) + \mathcal{D}(y_{(s+1):r}) > G(r)$, then for any $q > r$, it is mathematically impossible for the most recent changepoint before the q-th time point to have occurred at s. We prove this criterion formally below, which allows us to prune the set of candidate changepoints and speed up the DP algorithm empirically.

Algorithm 2 Active set DP Algorithm for the ℓ_0 problem

Input : Input fluorescence sequence $F_t, t \in [T]$, Parameters $\gamma \in (0,1)$, $\lambda > 0$
Output : Estimated calcium concentration $\hat{c}_t, t \in [T]$, estimated spikes $\hat{s}_t$
Initialize $cp(0) = \emptyset$, $G(0) = -\lambda$, $\mathcal{E}_0 = \{0\}$
for $t = 1, \cdots, T$ **do**
  Compute $G(t) = \min_{s \in \mathcal{E}_s} \{G(s) + \mathcal{D}(F_{s:t}; \gamma) + \lambda\}$
  Set $t' = \arg\min_{s \in \mathcal{E}_s} \{G(s) + \mathcal{D}(F_{s:t}; \gamma) + \lambda\}$
  Update $cp(t) = cp(t') \bigcup t'$
  Update $\mathcal{E}_{t+1} = \{s \in \mathcal{E}_t \cup t : G(s) + \mathcal{D}(F_{s:t}; \gamma) + \lambda < G(t)\}$
end
$\hat{s}_t = 1(t \in cp(T))$

$$\hat{c}_t = \begin{cases} \gamma \cdot \hat{c}_{t-1} & \hat{s}_t = 0 \\ \arg\min_y \sum_{t'=t}^{\min \tilde{t} > t : \hat{s}_{\tilde{t}} \neq 0} (F_{t'} - \gamma^{t'-t} \cdot y)^2 & \hat{s}_t \neq 0 \end{cases}$$

Corollary 2.1 (Correctness of Algorithm 2)**.** *Algorithm 2 returns the correct output in fewer operations*

Proof. Note that the pruning step in Algorithm 2 will discard s for all the future iterations if $\exists r > s$ s.t. $G(s) + \mathcal{D}(y_{s:t}) + \lambda < G(r)$. Then in order for this to yield a correct algorithm, it suffices to prove that for any time points $q > r$, we will always have
$G(s) + \mathcal{D}(F_{s:q}) + \lambda > G(r) + \mathcal{D}(F_{r:q}) + \lambda$, i.e., the $\arg\min$ is not changed by this pruning step.

Applying the pruning condition $G(s) + \mathcal{D}(y_{s:t}) + \lambda < G(r)$ and rearranging the equation implies that the desired statement is equivalent to proving that
$\mathcal{D}(F_{r:q}) + \mathcal{D}(F_{s:r}) < \mathcal{D}(F_{s:q})$. Now RHS, by definition, is

$$\min_c \frac{1}{2} \sum_{i=s}^{q} (F_i - \gamma^{i-s} c)^2 = \min_c \frac{1}{2} \sum_{i=s}^{r} (F_i - \gamma^{i-s} c)^2 + \frac{1}{2} \sum_{i=r}^{q} (F_i - \gamma^{i-s} c)^2$$

$$\leq \min_{c_1} \frac{1}{2} \sum_{i=s}^{r} (F_i - \gamma^{i-s} c_1)^2 + \min_{c_2} \frac{1}{2} \sum_{i=r}^{q} (F_i - \gamma^{i-r} c_2)^2 = \mathcal{D}(F_{s:r}) + \mathcal{D}(F_{r:q})$$

□

Under mild conditions one can prove that the running time of the pruning algorithm is asymptotically more efficient, in expectation, compared to the original algorithm. We verify this claim empirically using simulated data.

Evaluation metric

If we have access to the ground truth of when a spike has occurred, we can measure the goodness of fit using distance on the spike sequences.

In particular, we consider the Van Rossum distance and the Victor–Purpura distance:

- Van Rossum distance: the Van Rossum distance between two discrete spike trains X and Y are computed by transforming them into continuous functions by convolving each spike t_k with an exponential kernel, e.g., $x(t) = \exp(-\frac{t}{\tau_R})$ where τ_R is the time constant. Now the distance will be $D_R(\tau_R) = \sqrt{\frac{1}{\tau_R} \int_0^\infty [x(t) - y(t)]^2 dt}$.
- Victor–Purpura distance: Victor and Purpura defined the distance between two spike sequences in terms of the minimum cost of transforming one spike train into the other by means of just three basic operations: spike insertion, spike deletion, and shifting a spike by some interval Δt.

Interested readers can find more discussions on spike metrics in papers such as [Victor and Purpura, 1996, Chicharro et al., 2011, Houghton and Kreuz, 2012, van Rossum, 2001], but the main takeaway is that both define a metric which measures the distance between two spike sequences.

Hyper-parameter tuning

- λ: Since users would most likely not have access to a training set where the true occurrence of spikes is recorded, we propose a 2-fold cross-validation rule based on the goodness of fit of the calcium curves. On a high level, we train the algorithms with a grid of λ's on the odd data points of the fluorescent trace and pick the λ that gives the minimum mean squared error (MSE) on the even time points, where the predictions are generated by $\hat{c}_{2k} = \frac{1}{2}(\hat{c}_{2k-1} + \hat{c}_{2k+1}), \forall k \in \mathbb{N}$. We assume that readers are familiar with the basic ideas of cross-validation and hyper-parameter tuning and won't outline a detailed algorithm; however interested readers can refer to [Hastie et al., 2009] for more details.

- γ: while the decaying parameter γ can be in principle jointly estimated, as pointed out in [Pachitariu et al., 2018], estimating the decay rate is surprisingly difficult. There are two heuristics we provide: first, one can pick out a segment of data which exhibit the autoregressive decaying property visually and estimate γ based on the chosen segment. The second and the more popular heuristic is used by [Vogelstein et al., 2010] where the algorithm sets $\gamma = 1 - \frac{\phi}{\delta_t}$. ϕ is a time-scale parameter determined based on the calcium indicator molecules used in the experiment, and δ_t is the time frame of recording. Throughout our simulation we assume that γ is given but for recorded fluorescent traces we adopt the second method. We also included additional simulations to test the sensitivity of our results to mis-specifications of γ.

Implementation

We implemented the algorithm in `Python` (version 3.6.1) and used the just-in-time compiler `Numba` for speeding up the key computation step. Codes for reproducing our results are hosted on `GitHub` with a Jupyter notebook demonstrating the use of the proposed algorithm.

Result

Timing result

The data were simulated with $\gamma = 0.998, \sigma = 0.15$ and true spikes $s_t \overset{\text{ind}}{\sim} Poisson(\theta)$ with $\theta \in \{0.1, 0.01, 0.001\}$. We solved the ℓ_0 problem with $\lambda = 1$ with both Algorithms 1 and 2. Timing results, averaged over 50 iterations, were plotted in Figure 3. As expected the running time of Algorithm 1 scales quadratically in the length of the traces, whereas the running time of Algorithm 2 is upper bounded by that of Algorithm 1 and decreases as the firing rate increases. Using 2 and our C-extended `Python` code, we are able to solve the ℓ_0 problem in less than a minute for recorded sequences of length 100,000 with moderate ($\theta \approx 0.01$) firing rate on a MacBook Pro laptop. Note that most of the recorded fluorescence traces are of similar if not shorter length and therefore the active pruning algorithm indeed offers a practical spike inference procedure. Similar to the original authors' finding, we note that our Algorithm 2 is still slower than the algorithm in [Friedrich et al., 2017] for solving the ℓ_1 problem, which is implemented in `Cython` and has approximately linear time. While it's beyond the scope of this paper, a dramatic speed-up of the ℓ_0

algorithm was proposed and tested in a follow-up paper by the authors and collaborators(See [Jewell et al., 2019]) , exhibiting competitive time complexity to that of [Friedrich et al., 2017].

Results on simulated datasets

In this section, we use *in silico* data to demonstrate the performance advantages of the proposed ℓ_0 problem over the ℓ_1 approaches():

1. The ℓ_1 estimator proposed in [Friedrich et al., 2017], which involves a single tuning parameter λ penalizing the magnitude of spikes (analogous to our λ).

2. A thresholded version of the ℓ_1 estimator: given a choice of λ and suppose solutions to the ℓ_1 problem $\hat{s}_t$, the post-thresholdiung estimator is defined to be $\tilde{s}_{t,L} = \hat{s}_t 1(\hat{s}_t \geq L), \forall t \in [T]$ for some positive constant L. In other words, the post-thresholding estimation procedure only retains the estimated spikes whose increase in calcium exceeds a thresdhol L. The post-thresholding estimator therefore involves two tuning parameter L and λ, and is primarily motivated by the performance increase in real dataset as noted in [Friedrich et al., 2017]. We remark that the same observation applies in our toy example (Figure 2) where in the last (rightmost) panel, the post-thresholded (orange) estimators have superb performance compared to its un-thresholded counterpart.

As we have access to the ground truth about spike times, we measure our performance of each method based on two criteria: 1. error in calcium concentration estimation measured by MSE and 2. error in spike detection using the Victor–Purpura distance metric and van Rossum distance (use $\tau_R = 2$) introduced in the method section.

We generated 50 simulated datasets according to our data generating process with parameters $\gamma = 0.96, T = 1000, \sigma = 0.15$ and $s_t \overset{i.i.d.}{\sim} Poisson(0.01)$. On each simulated dataset we solved the ℓ_0 and ℓ_1 problem for a range of values of the tuning parameter λ. Moreover, we post-thresholded the ℓ_1 solution, with five different threshold values: $L \in \{0, 0.125, 0.250, 0.375, 0.500\}$. Figure 4 displays the averaged results of our simulation: from top to bottom, the first panel plots the MSE of estimated calcium as a function of λ; the second panel plotted the Victor–Purpura distance between estimated and true spikes s_t (log scale) as a function of λ; the third panel in spike event detection for the van Rossum distance distance between estimated and true spikes s_t (log scale).

Figure 4 reveals that, if we specify γ correctly, ℓ_0 estimator results in dramatically lower errors in both calcium estimation and spike detection than the ℓ_1 estimator. Note that while choosing $L = 0.5$ gives us drastically lower error, the the post-thresholding procedure involves two tuning parameters, λ and L, which leads to challenges in tuning parameter selection. In addition, we note that the λ value leading to the minimal MSE in calcium concentration matches up with the λ leading to the minimal spike distance errors, therefore suggesting the plausibility of using a cross-validation scheme to select λ as outlined in the method section. In contrast, there is a discrepancy between the λ's that give the minimal calcium estimation error and the spike estimation error for the ℓ_1 estimators. Intuitively, this is a consequence of the ℓ_1 penalty simultaneously reducing the number of estimated spikes and shrinks the estimated calcium concentration. This means that the same cross-validation scheme will not perform well for the ℓ_1 approach.

Results on real datasets

In this section we apply the proposed ℓ_0 approach and the ℓ_1 approach in [Friedrich et al., 2017] (both with and without post-thresholding), to two calcium imaging datasets. In the first dataset, the true spike times were measured using electrophysiology sisimultaneously

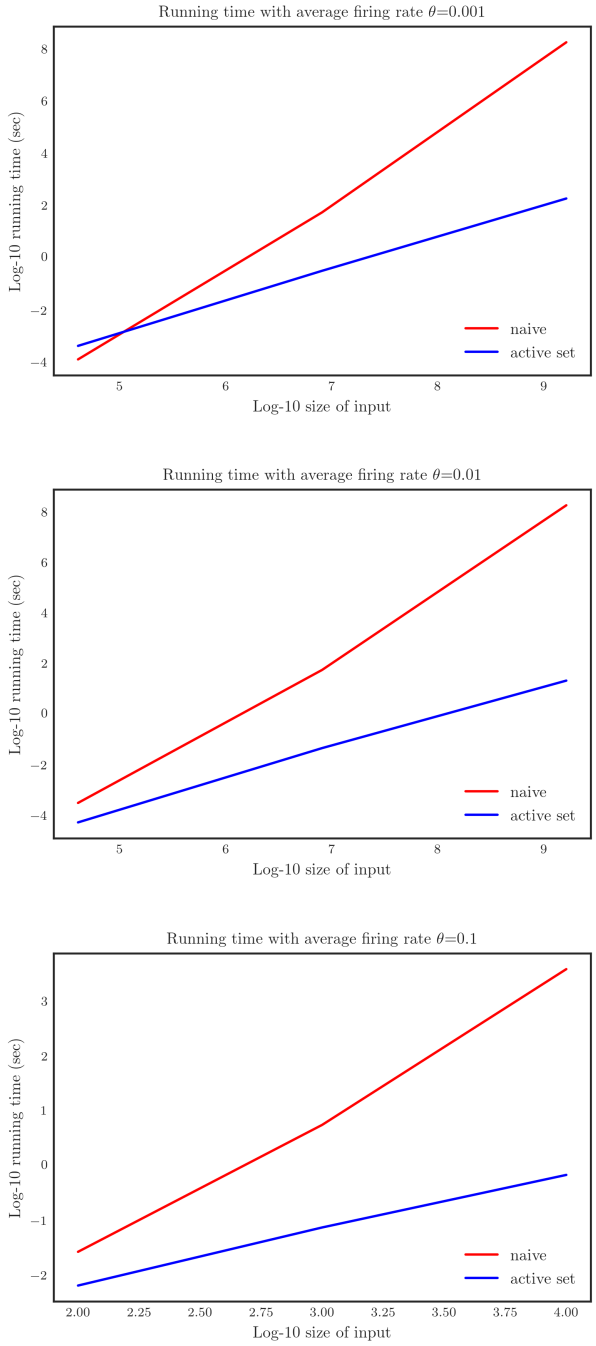

Figure 3. Timing results for solving the ℓ_0 problem for the global optimum, using Algorithms 1 (Naive) and 2 (Active set). The x-axis displays the length of the input sequence in log10-scale, and the y-axis displays the average running time in seconds. Each panel, from top to bottom, corresponds to data simulated according to our data generating process with $s_t \overset{i.i.d.}{\sim} Poisson(\theta)$, with $\theta = 0.001, 0.01, 0.1$ respectively. Standard errors over 50 repetitions are on average $< 1\%$ of the mean compute time and therefore not plotted.

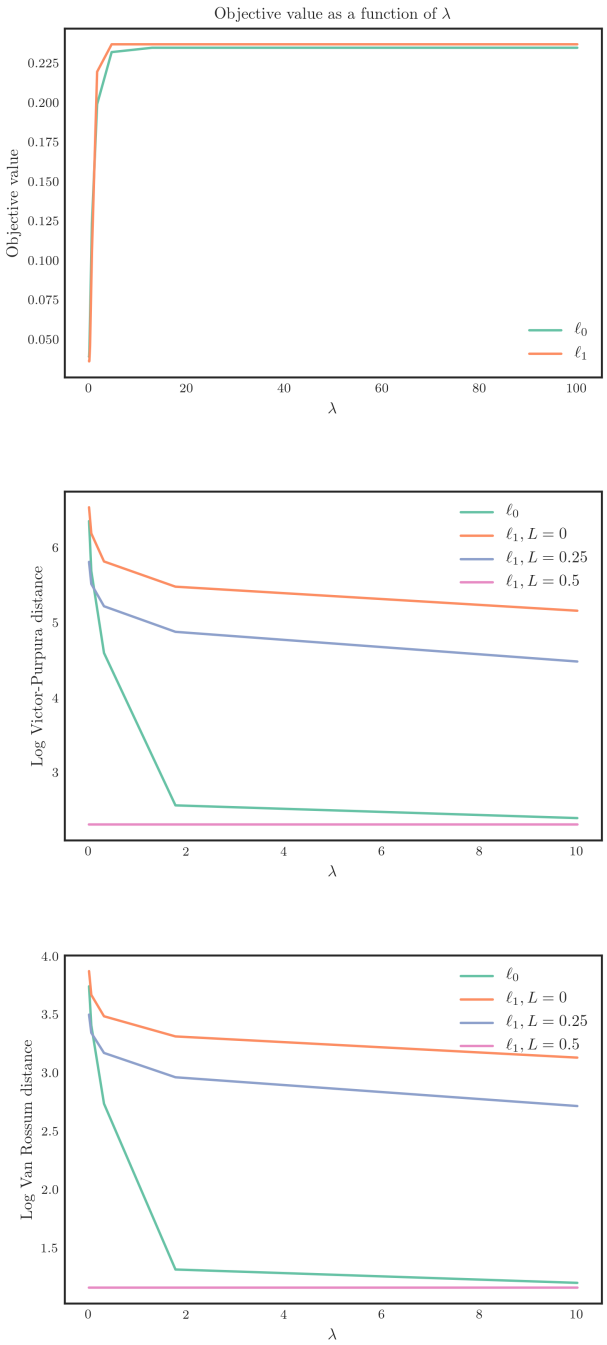

Figure 4. Simulation study to assess the error in spike detection and calcium estimation using the ℓ_0, ℓ_1, and post-thresholded ℓ_1 estimators. From top to bottom, first panel: Error in estimated calcium estimation measured by MSE; second panel: Error in spike detection, measured using Victor-Purpura distance; third panel: Error in spike detection measured using van Rossum distance.

and therefore we can directly assess the spike detection accuracy of each proposal. In the second dataset the true spike times are unknown [**?**, Sunkin et al., 2013]; nonetheless, we

will make a qualitative comparison of the results of the ℓ_1 and ℓ_0 proposals.

Application to [Chen et al., 2013] data

We first consider a dataset recorded in [Chen et al., 2013] that consists of simultaneous calcium imaging and electrophysiological measurements, obtained from the Collaborative Research in Computational Neuroscience portal (http://crcns.org/data-sets/methods/cai-1/about-cai-1). In what follows we refer to the spike times obtained from the electrophysiological measurements as the "true" spikes. Figure 5 shows a selected 40-second recording from cell 2002 which expresses the calcium indicator GCaMP6. The data were measured at 60 Hz for a total of 2400 time points. The raw fluoresence traces are DF/F transformed with a 20% percentile filter as documented in [Friedrich et al., 2017]. There are in total 23 true spikes in the 40-second recording and we estimated the γ to be 0.986 using the first few time steps, which turns out to be really close to the value obtained from the decaying parameters of GCaMP6. The top panel of Figure 4 shows a 40-second recording from cell 2002, which expresses GCaMP6s. The data are measured at 60 Hz for a total of 2400 time steps. While in [Jewell and Witten, 2018] the authors chose λ for the ℓ_0 problem such that 23 spikes are spiked, we consider such approach *ex post facto* and inconsisitent with the cross-validation approach proposed in the paper. Therefore we use a grid search to obtain the optimal values of λ for both ℓ_0 and ℓ_1 problems. We also consider the post-thresdholded estimators with $L = 0, 0.125, 0.25$; however we remind the readers that L is an additional parameter which needs to be chosen rigorously in practice.

From Figure 5, we see that ℓ_0 estimators perform competitively against all ℓ_1 counterparts, thresholded or not. In particular, there are only few false positives where the ℓ_1 problems, even after thresholding, are not able to recover some true spikes without overestimating the total amount of spikes drastically, especially in short succession.

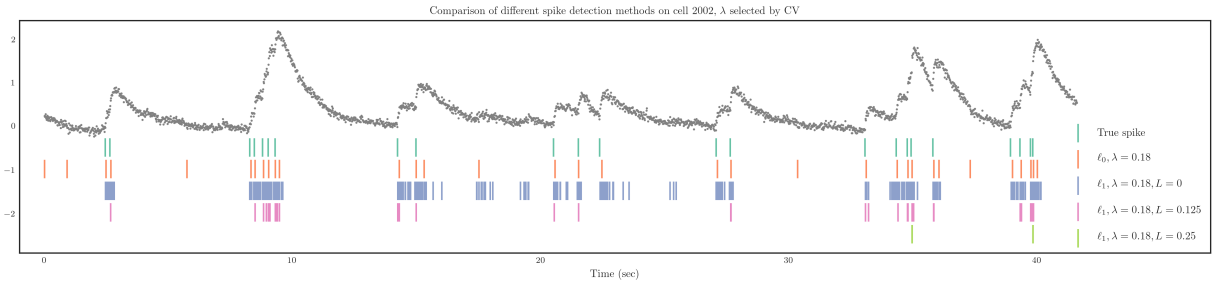

Figure 5. Spike detection result for cell 2002 of the [Chen et al., 2013] data. The observed fluorescence are plotted in grey dots and true spikes are plotted in verticle bars in teal. Estimated spike times from the ℓ_0 problem are shown in orange and the estimated spike times from the ℓ_1 problem with different post-thresholding values (0, 0.125, 0.25) are shown in livid, pink, and lime respectively).

We also note that the ℓ_0 method tends to estimate spike times one or two time steps ahead of the true spike times. This can partially due to model mis-specification: assuming a first-order autoregressive trades off the computational tractability with model generalizability. For instance, we can see from the true recorded traces in Figure 5 that a spike event is followed by an increase in calcium over the course of a few timesteps before the onset of the exponential decay phase. We discuss a few possible avenues to address this minor issue in the discussion section.

Application to Allen brain observatory data

We now consider a dataset from the Allen Brain Observatory, a large open-source repository of calcium imaging data from the mouse visual cortex [Allen-Institute,]. Like most calcium imaging experiments, this dataset does not contain the true spike times, so it is difficult to objectively assess the performances of different algorithms. Instead, as a proof of concept, we fit two few different sets of λ values where one of them is picked using the cross-validation scheme and another one is picked so that qualitatively fewer spikes are estimated. We then compare the results qualitatively using visual inspection.

In particular, we consider the second cell in the experiment 510221121 of the Allen Brain Observatory data and applied the proposed approaches for the first 5,000 time steps of the DF/F-transformed fluorescence traces. Since the data are measured at 30 Hz, this amounts to the first 100 seconds of the recording.

We used $\gamma = 0.982$ throughout the experiment and Figure 6 shows the ℓ_0, ℓ_1, and post-thresholded ℓ_1 estimators with λ's chosen using cross-validation on the calcium estimation. As in the previous subsection on the data from [Chen et al., 2013], we see that, when faced with a really large increase in fluorescence such as the wave of spikes from 30-35 seconds, the ℓ_1 estimators (without no thresholding, i.e., $L = 0$) tend to estimate a large number of spikes concentrated in a few time points. However we do see that with the number of estimated spikes on the order around 10-100, even the ℓ_0 estimators have very dense estimates, which motivates us to look at the behavior of different estimators with larger penalties.

In Figure 7, we see that with much fewer total number of spikes estimated, the advantage of ℓ_0 is even more pronounced: the ℓ_0 estimators seem to have put a spike everywhere with a significant increase in the calcium concentration where the ℓ_1 estimators, even after post-thresholding, are not placing the spikes even remotely correctly.

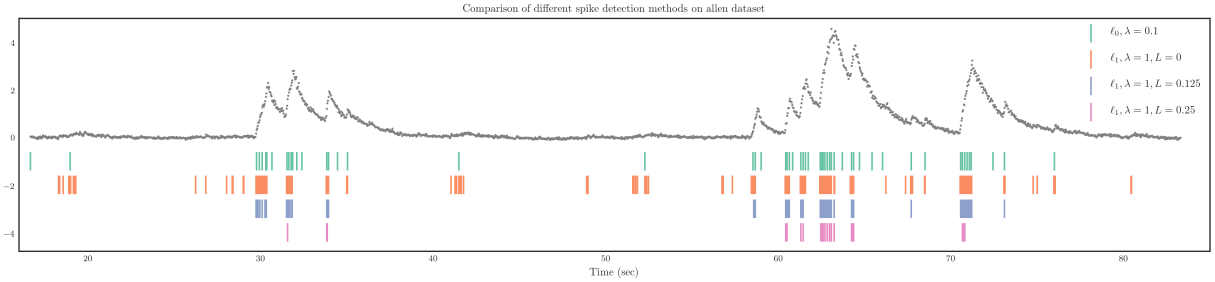

Figure 6. Comparison of estimated spikes from different approaches with cross-validated chosen λ: The first 5,000 time steps from the second cell in exepriment 510221121 from the Allen Brain Observatory. The DF/F-transformed fluorescence is plotted using gray dots and the estimated spikes from the ℓ_0 problem are plotted in teal; the estimated spikes from the ℓ_1 problem with $L = 0, 0.125, 0.25$ are plotted in orange, livid, and pink respectively.

Upon visual inspection with different λ choices, though we don't have access to the true spike times, the results from the ℓ_0 proposal seem superior to those of the ℓ_1 and post-thresholded ℓ_1 proposals.

Discussion

The work summarized above ([Jewell and Witten, 2018]) leads to a few future directions for exploration.

Firstly, we recall the issue that the AR-1 assumption doesn't always hold in reality and yields ℓ_0 estimators which appear to be "delayed". This can be addressed by either

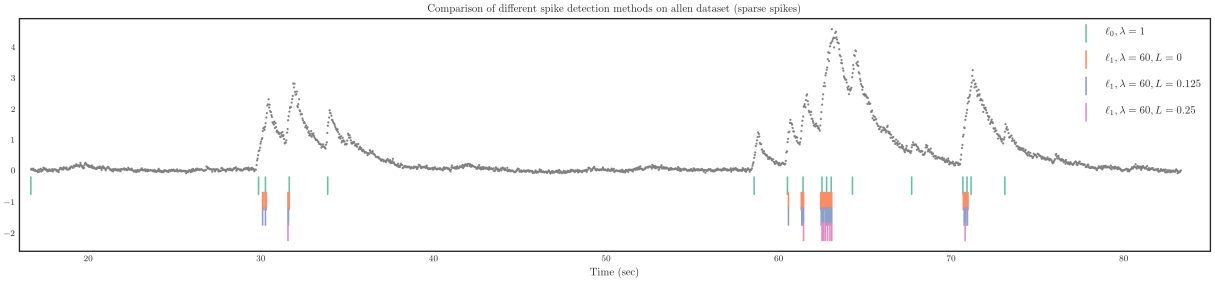

Figure 7. Comparison of estimated spikes from different approaches with λ such that around 20 spikes were estimated for ℓ_0 and ℓ_1 (no post-thresholding) estimators: The first 5,000 time steps from the second cell in exepriment 510221121 from the Allen Brain Observatory. The DF/F-transformed fluorescence is plotted using gray dots and the estimated spikes from the ℓ_0 problem are plotted in teal; the estimated spikes from the ℓ_1 problem with $L = 0, 0.125, 0.25$ are plotted in orange, livid, and pink respectively.

empirically adjusting the estimates to align with the true spike times, provided that the ground truth is available, or considering a more elaborate AR-p process based changepoint detection process. In particular, we note that the OASIS algorithm for solving the ℓ_1 problem in [Friedrich et al., 2017] can accommodate $p = 2$ readily. While [Jewell and Witten, 2018] claimed an extension for the AR-p process should be straightforward, this is still active work in progress.

Secondly, we consider the speed of the algorithm and the dropped $c_t - \gamma c_{t-1} \geq 0$ constraint. The violation of the latter constraint is often referred to as the "negative spike" problem and is biologically implausible. Both of these problems have been addressed by the follow-up work in [Jewell et al., 2019] which uses a functional pruning algorithm, resulting in drastic speed-up and removal of the "negative spikes". While an in-depth implementation and discussion of [Jewell et al., 2019] is out of the scope of our review, we encourage interested readers to read their papers for more insights and to use the corresponding software provided in [Jewell et al., 2019].

Finally we consider the lingering question of theoretical justification and analysis of our empirical observations. It's been long known that the ℓ_0 and ℓ_1 norms, together with a family of induced optimization problems, are closely related. In particular, in the special case of $\gamma = 1$, the ℓ_1 problem outlined in this expository can be as a fused lasso optimization problem, which has been extensively studied. Some of the selected works include [Qian and Jia, 2016, Zou, 2006, Harchaoui and Lévy-Leduc, 2010, Owrang et al., 2017, Bleakley and Vert, 2011, **?**, Hastie et al., 2015]. However, the authors have left the theoretical analysis such as the ℓ_2 error bounds of the estimated calcium concentration and the spike recovery properties to future work. In particular, this paper focused on the empirical comparison of the ℓ_0 and ℓ_1 solutions since there is no guarantee that approximating the ℓ_0 norm with ℓ_1 norm will yield an attractive local optimum on a given dataset.

Summary and critique

In this expository, we summarized and reviewed the paper by Jewell and Witten [Jewell and Witten, 2018] where we considered solving a seemingly intractable ℓ_0 optimization problem motivated by a neuroscience application. We showed that by recasting the optimization problem into the changepoint detection framework, an efficient dynamic programming algorithm can in fact solve for the global minimum. To evaluate the practical value of the proposed algorithm, we considered comparing and contrasting the performance the solutions to the ℓ_0 problem with the solutions to the ℓ_1 problem, which is often referred to as the

state-of-the-art technique on spike detection, on both simulated and real recordings of neuron activities. Empirically the ℓ_0 solutions achieve superior performance in both simulated and real datasets; thus suggesting a promising avenue for applying this algorithm to large scale recordings by neuroscientists. Indeed we see from the authors' follow-up paper [Jewell et al., 2019], the improved version of what we presented in this expository was applied to thousands of hours of recordings in the Allen Institute, advancing our understanding of neuroscience and brain connectivity.

While we congratulate the authors on a well motivated, efficiently implemented, and elegantly written paper, we do offer a few pieces of unsolicited suggestions below.

On methods and model formulation:

Clear model formulation and method proposition are the strength of [Jewell and Witten, 2018]. In particular, the autoregressive process based data generating process is easy to understand and well-motivated scientifically; one possible extension is probably to contrast the class of methods to some of the supervised machine learning based algorithms which attempt to learn a mapping from the calcium concentration to a binary sequence representing the presence of a spike [Theis et al., 2016, Berens et al., 2018].

In terms of model mis-specification discussion, nevertheless, we note that the authors did not experiment with AR-p data generating process with $p > 1$. For instance understanding how ℓ_0 estimators perform when $p = 2$ would be of practical interest, especially when the competing algorithm OASIS accommodates $p = 2$ as one of its options.

On simulations and numerical results

We notice a few inconsistencies and possible avenues of improvements in terms of the simulation and real data analysis performed by the authors.

- Presentation of simulated results: the authors presented the equivalent plot of our Figure 4 with the number of estimated spikes on the x-axis. While this provides great intuition on how the choice of λ affects the number of estimated spike, in practice one often has no knowledge of the number of actual spikes. Therefore the presentation of results in the original paper focused heavily on how well ℓ_0 performed compared to ℓ_1 given that both approaches estimated around the same number of spikes, which could be confusing for some readers, letting them believe that one needs to know the true number of spikes to use the ℓ_0 algorithm. We instead plotted results as a function of λ and at least in simulation it's clear that the cross-validation scheme for choosing λ would yield a good result in terms of both spike distance metrics

- Selection of λ on real datasets: as commented above, the authors are choosing λ based on the resulting number of estimated spikes. We find it more natural to present results based on the proposed cross-validation scheme for selecting λ. However we do find the authors' approach illuminating on the Allen Institute data where the true spike times are unavailable.

- Specification of γ: little is discussed on the behavior of the solutions to ℓ_0 and ℓ_1 problems when the decaying parameter γ is misspecified. While this is of secondary interest, we find it important in practice when a reliable estimation of γ might not be readily available. We included the details of simulation results in the Appendix where we consider the case where we overestimated the parameter, i.e., $\hat{\gamma} = 0.99$ while $\gamma = 0.96$ and where we underestimated the parameter, i.e., $\hat{\gamma} = 0.96$ while $\gamma = 0.99$. Based on these two simple simulation settings, we find ℓ_0-based estimators appear to be more sensitive to the mis-specification of γ. In other words when we have large estimation errors in γ, the advantages of ℓ_0 relative to ℓ_1 seem to have disappeared.

We conjecture that the sensitivity is due to the discrete nature of the objective function in the ℓ_0 problem, but empirically a 0.03 difference would probably be drastic.

On technical proofs

Since [Jewell and Witten, 2018] focused on the empirical aspect of the spike inference algorithms, there is little to discuss in terms of the proof details. We provide the addition MAP motivation in the appendix to motivate different optimization objectives.

Appendix

Glossary of notations

- $[T] = \{1, 2, \cdots, T\}$ for $T \in \mathbb{N}$
- s_t: spike at time point t
- c_t: calcium concentration at time point t
- F_t: observed fluorescence at time point t
- $O(f(x))$: a function $g(x)$ is $O(f(x))$ if $\lim_{x\to\infty} \frac{g(x)}{f(x)} = 0$. In particular, for time complexity analysis we often consider non-negative functions.
- $1(A)$: indicator function, $= 1$ if A occurs and 0 otherwise
- ℓ_0 problem:

$$\begin{aligned} \min_{c_1,\cdots,c_T,s_1,\cdots,s_T\in\mathbf{R}} \quad & \sum_{t=1}^{T}(F_t - c_t)^2 + \lambda 1(s_t \neq 0) \\ \text{subject to} \quad & s_t = c_t - \gamma c_{t-1}, \forall t \in [T] \end{aligned} \tag{6}$$

  and we refer to the solutions $\hat{s}_t$ as the ℓ_0 solutions.

- ℓ_1 problem:

$$\begin{aligned} \min_{c_1,\cdots,c_T,s_1,\cdots,s_T\in\mathbf{R}} \quad & \sum_{t=1}^{T}(F_t - c_t)^2 + \lambda|s_t| \\ \text{subject to} \quad & s_t = c_t - \gamma c_{t-1}, \forall t \in [T] \\ & c_t - \gamma c_{t-1} \geq 0, \forall t \in [T] \end{aligned}$$

  and we refer to the solutions $\hat{s}_t$ as the ℓ_1 solutions.

Maximum A Posteriori estimation

In this section, we motivate different optimization objectives using the principle of Maximum A Posteriori estimation. Consider the same data generating process described in Figure 1, we first consider to estimate s_t given F_t, with a Poisson prior on s_t.

$$\hat{s}_t = \text{argmax}_{s_t\in\mathbb{N}} P(s_1, \cdots, s_T|F)$$

$$= \text{argmax}_{s_t\in\mathbb{N}, c_{1:T}} \sum_{t=1}^{T} -\frac{1}{2\sigma^2}(F_t - c_t)^2 + s_t \log\theta - \log s_t!$$

. However we note that this optimization problem over $s_t \in \mathbb{N}$, as noted by [Vogelstein et al., 2010], is highly non-convex and computationally intractable.

Now we present a continuous relaxation of the estimation above where we put an expoential prior on s_t:

$$\hat{s}_t = \text{argmax}_{s_t>0} \sum_{t=1}^{T} -\frac{1}{2\sigma^2}(F_t - c_t)^2 - s_t \cdot \theta$$

$$\overset{equiv}{=} \text{argmax}_{c_1,\cdots,c_T; c_t - \gamma c_{t-1} > 0} \sum_{t=1}^{T} -\frac{1}{2\sigma^2}(F_t - c_t)^2 + (c_t - \gamma c_{t-1})$$

$$\approx \text{argmin}_{c_1,\cdots,c_T} \frac{1}{2} \sum_{t=1}^{T} (F_t - c_t)^2 + \lambda |c_t - \gamma c_{t-1}|$$

where in the last step we used the constraint to guarantee the non-negativeness of our estimated spikes.

Finally we consider the binarized Poisson random variable approach in [Jewell and Witten, 2018] which gives us the following optimization problem

$$\hat{s}_t = \text{argmax}_{s_t \in \{0,1\}} P(s_1, \cdots, s_T | F)$$

$$= \text{argmax}_{s_t \in \{0,1\}} \sum_{t=1}^{T} -\frac{1}{2\sigma^2}(F_t - c_t)^2 + s_t \log\theta - \log s_t!$$

$$= \text{argmin}_{s_t \in \{0,1\}} \frac{1}{2} \sum_{t=1}^{T} (F_t - c_t)^2 + \lambda \cdot s_t$$

.

Sensitivity analysis of the specification of γ

In this section we present the result anaglous to those in Figure 4 with mis-specified γ; we can see that performance of both calcium estimation and spike estimation of the ℓ_0 estimators decreased drastically compared to those of the ℓ_1 estimators, indicating that the behavior of the ℓ_0 estimators might be more sensitive to the value of γ.

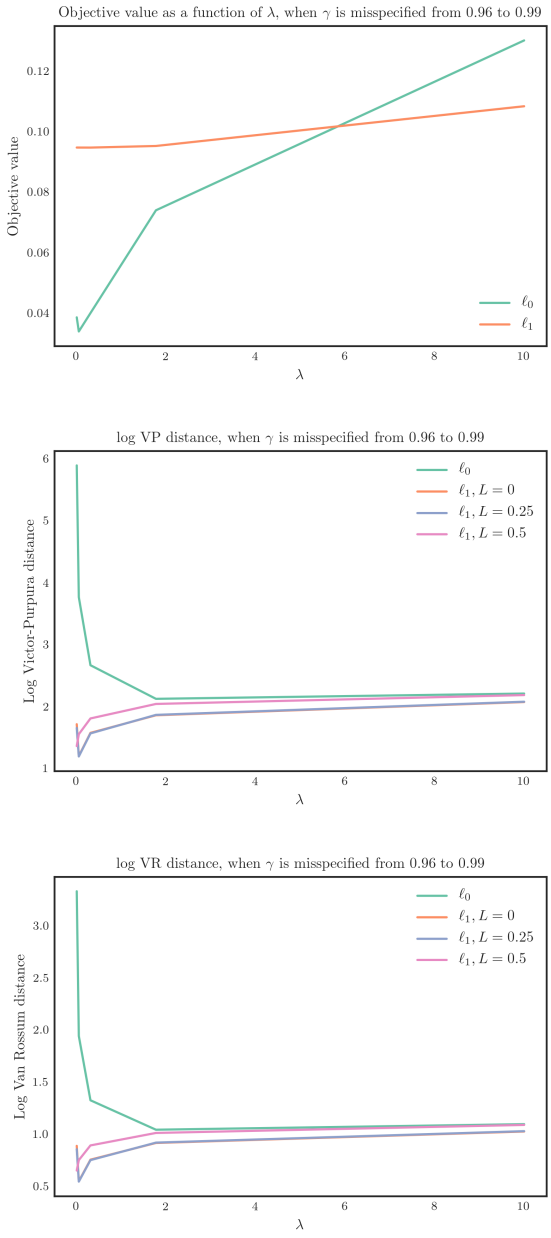

Figure 8. Simulation study to assess the error in spike detection and calcium estimation using the ℓ_0, ℓ_1, and post-thresholded ℓ_1 estimators. From top to bottom, first panel: Error in estimated calcium estimation measured by MSE; second panel: Error in spike detection, measured using Victor-Purpura distance; third panel: Error in spike detection measured using van Rossum distance.

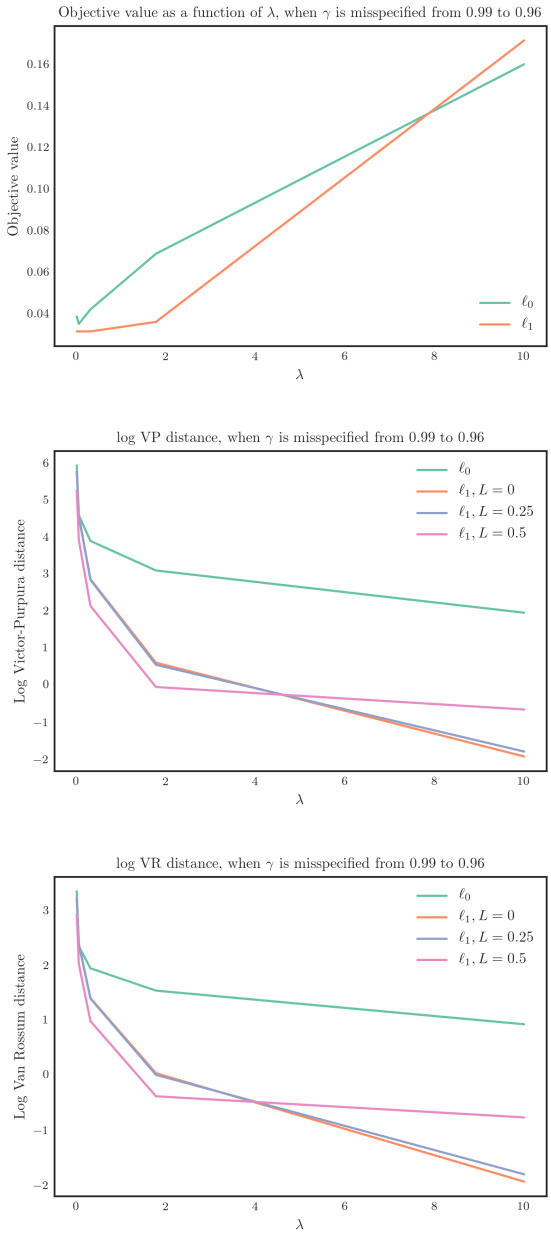

Figure 9. Simulation study to assess the error in spike detection and calcium estimation using the ℓ_0, ℓ_1, and post-thresholded ℓ_1 estimators. From top to bottom, first panel: Error in estimated calcium estimation measured by MSE; second panel: Error in spike detection, measured using Victor-Purpura distance; third panel: Error in spike detection measured using van Rossum distance.

References

Allen-Institute, . Allen-Institute. Allen brain observatory. http://observatory.brain-map.org/.

Auger and Lawrence, 1989. Auger, I. E. and Lawrence, C. E. (1989). Algorithms for the optimal identification of segment neighborhoods. *Bull. Math. Biol.*, 51(1):39–54.

Berens et al., 2018. Berens, P., Freeman, J., Deneux, T., Chenkov, N., McColgan, T., Speiser, A., Macke, J. H., Turaga, S. C., Mineault, P., Rupprecht, P., Gerhard, S., Friedrich, R. W., Friedrich, J., Paninski, L., Pachitariu, M., Harris, K. D., Bolte, B., Machado, T. A., Ringach, D., Stone, J., Rogerson, L. E., Sofroniew, N. J., Reimer, J., Froudarakis, E., Euler, T., Román Rosón, M., Theis, L., Tolias, A. S., and Bethge, M. (2018). Community-based benchmarking improves spike rate inference from two-photon calcium imaging data. *PLoS Comput. Biol.*, 14(5):e1006157.

Bleakley and Vert, 2011. Bleakley, K. and Vert, J.-P. (2011). The group fused lasso for multiple change-point detection.

Boyd and Vandenberghe, 2004. Boyd, S. and Vandenberghe, L. (2004). Convex optimization — optimization, OR and risk. https://www.cambridge.org/gb/academic/subjects/statistics-probability/optimization-or-and-risk/convex-optimization?format=HB. Accessed: 2019-6-2.

Carsen et al., 2018. Carsen, Stringer, Marius, and Pachitariu (2018). Computational processing of neural recordings from calcium imaging data.

Chen et al., 2013. Chen, T.-W., Wardill, T. J., Sun, Y., Pulver, S. R., Renninger, S. L., Baohan, A., Schreiter, E. R., Kerr, R. A., Orger, M. B., Jayaraman, V., Looger, L. L., Svoboda, K., and Kim, D. S. (2013). Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature*, 499(7458):295–300.

Chicharro et al., 2011. Chicharro, D., Kreuz, T., and Andrzejak, R. G. (2011). What can spike train distances tell us about the neural code? *J. Neurosci. Methods*, 199(1):146–165.

Dombeck et al., 2007. Dombeck, D. A., Khabbaz, A. N., Collman, F., Adelman, T. L., and Tank, D. W. (2007). Imaging large-scale neural activity with cellular resolution in awake, mobile mice. *Neuron*, 56(1):43–57.

Friedrich et al., 2017. Friedrich, J., Zhou, P., and Paninski, L. (2017). Fast online deconvolution of calcium imaging data. *PLoS Comput. Biol.*, 13(3):e1005423.

Glimcher, 2011. Glimcher, P. W. (2011). Understanding dopamine and reinforcement learning: the dopamine reward prediction error hypothesis. *Proc. Natl. Acad. Sci. U. S. A.*, 108 Suppl 3:15647–15654.

Harchaoui and Lévy-Leduc, 2010. Harchaoui, Z. and Lévy-Leduc, C. (2010). Multiple Change-Point estimation with a total variation penalty. *J. Am. Stat. Assoc.*, 105(492):1480–1493.

Hastie et al., 2009. Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*. Springer Series in Statistics. Springer-Verlag New York, 2 edition.

Hastie et al., 2015. Hastie, T., Tibshirani, R., and Wainwright, M. (2015). *Statistical Learning with Sparsity: The Lasso and Generalizations*. Chapman & Hall/CRC.

Houghton and Kreuz, 2012. Houghton, C. and Kreuz, T. (2012). On the efficient calculation of van rossum distances. *Network*, 23(1-2):48–58.

Jackson et al., 2003. Jackson, B., Scargle, J. D., Barnes, D., Arabhi, S., Alt, A., Gioumousis, P., Gwin, E., Sangtrakulcharoen, P., Tan, L., and Tsai, T. T. (2003). An algorithm for optimal partitioning of data on an interval.

Jewell and Witten, 2018. Jewell, S. and Witten, D. (2018). EXACT SPIKE TRAIN INFERENCE VIA 0 OPTIMIZATION. *Ann. Appl. Stat.*, 12(4):2457–2482.

Jewell et al., 2019. Jewell, S. W., Hocking, T. D., Fearnhead, P., and Witten, D. M. (2019). Fast nonconvex deconvolution of calcium imaging data. *Biostatistics*.

Killick et al., 2012. Killick, R., Fearnhead, P., and Eckley, I. A. (2012). Optimal detection of changepoints with a linear computational cost. *J. Am. Stat. Assoc.*, 107(500):1590–1598.

Mukamel et al., 2009. Mukamel, E. A., Nimmerjahn, A., and Schnitzer, M. J. (2009). Automated analysis of cellular signals from large-scale calcium imaging data. *Neuron*, 63(6):747–760.

Owrang et al., 2017. Owrang, A., Malek-Mohammadi, M., Proutiere, A., and Jansson, M. (2017). Consistent change point detection for piecewise constant signals with normalized fused LASSO. *IEEE Signal Process. Lett.*, 24(6):799–803.

Pachitariu et al., 2018. Pachitariu, M., Stringer, C., and Harris, K. D. (2018). Robustness of spike deconvolution for neuronal calcium imaging. *J. Neurosci.*, 38(37):7976–7985.

Paninski and Cunningham, 2018. Paninski, L. and Cunningham, J. P. (2018). Neural data science: accelerating the experiment-analysis-theory cycle in large-scale neuroscience. *Curr. Opin. Neurobiol.*, 50:232–241.

Paninski et al., 2007. Paninski, L., Pillow, J., and Lewi, J. (2007). Statistical models for neural encoding, decoding, and optimal stimulus design. *Prog. Brain Res.*, 165:493–507.

Pnevmatikakis, 2018. Pnevmatikakis, E. A. (2018). Analysis pipelines for calcium imaging data. *Curr. Opin. Neurobiol.*, 55:15–21.

Qian and Jia, 2016. Qian, J. and Jia, J. (2016). On stepwise pattern recovery of the fused lasso. *Comput. Stat. Data Anal.*, 94:221–237.

Reid, 2012. Reid, R. C. (2012). From functional architecture to functional connectomics. *Neuron*, 75(2):209–217.

Smetters et al., 1999. Smetters, D., Majewska, A., and Yuste, R. (1999). Detecting action potentials in neuronal populations with calcium imaging. *Methods*, 18(2):215–221.

Stosiek et al., 2003. Stosiek, C., Garaschuk, O., Holthoff, K., and Konnerth, A. (2003). In vivo two-photon calcium imaging of neuronal networks. *Proc. Natl. Acad. Sci. U. S. A.*, 100(12):7319–7324.

Stringer and Pachitariu, 2018. Stringer, C. and Pachitariu, M. (2018). Computational processing of neural recordings from calcium imaging data. *Curr. Opin. Neurobiol.*, 55:22–31.

Sunkin et al., 2013. Sunkin, S. M., Ng, L., Lau, C., Dolbeare, T., Gilbert, T. L., Thompson, C. L., Hawrylycz, M., and Dang, C. (2013). Allen brain atlas: an integrated spatio-temporal portal for exploring the central nervous system. *Nucleic Acids Res.*, 41(Database issue):D996–D1008.

Theis et al., 2016. Theis, L., Berens, P., Froudarakis, E., Reimer, J., Román Rosón, M., Baden, T., Euler, T., Tolias, A. S., and Bethge, M. (2016). Benchmarking spike rate inference in population calcium imaging. *Neuron*, 90(3):471–482.

Tian et al., 2009. Tian, L., Hires, S. A., Mao, T., Huber, D., Chiappe, M. E., Chalasani, S. H., Petreanu, L., Akerboom, J., McKinney, S. A., Schreiter, E. R., Bargmann, C. I., Jayaraman, V., Svoboda, K., and Looger, L. L. (2009). Imaging neural activity in worms, flies and mice with improved GCaMP calcium indicators. *Nat. Methods*, 6(12):875–881.

van Rossum, 2001. van Rossum, M. C. (2001). A novel spike distance. *Neural Comput.*, 13(4):751–763.

Victor and Purpura, 1996. Victor, J. D. and Purpura, K. P. (1996). Nature and precision of temporal coding in visual cortex: a metric-space analysis. *J. Neurophysiol.*, 76(2):1310–1326.

Vogelstein et al., 2010. Vogelstein, J. T., Packer, A. M., Machado, T. A., Sippy, T., Babadi, B., Yuste, R., and Paninski, L. (2010). Fast nonnegative deconvolution for spike train inference from population calcium imaging. *J. Neurophysiol.*, 104(6):3691–3704.

Zeng and Sanes, 2017. Zeng, H. and Sanes, J. R. (2017). Neuronal cell-type classification: challenges, opportunities and the path forward. *Nat. Rev. Neurosci.*, 18(9):530–546.

Zou, 2006. Zou, H. (2006). The adaptive lasso and its oracle properties. *J. Am. Stat. Assoc.*, 101(476):1418–1429.