Inferring Neural Population Spiking Rate from Wide-Field Calcium Imaging

Merav Stern, Eric Shea-Brown, Daniela Witten

February 1, 2020

Abstract

Wide-field calcium imaging techniques allow recordings of high resolution neuronal activity across one or multiple brain regions. However, since the recordings capture light emission generated by the fluorescence of the calcium indicator, the neural activity that drives the calcium changes is masked by the calcium indicator dynamics. Here we develop and test novel methods to deconvolve the calcium traces and reveal the underlying neural spiking rate. Our methods take into account both the noise existent in the recordings and the temporal dynamics of the calcium indicator response. Our first method retrieves firing rates that are constant over discrete time bins. The size of each time bin depends on the data and is determined dynamically. Our second method retrieves the rate as a continuous function and is meant for studies that look for slow rate fluctuations rather than abrupt changes. We compare our results with those of two alternative approaches: direct deconvolution using a 'first differences' approach, and the 'Lucy-Richardson' image recovery method, adapted to recover temporal dynamics. We show that our methods outperform competitors on synthetic data as well as on wide-field calcium recordings in which parallel spiking recording was performed.

1 Introduction

Recent developments in optogenetics allow for the recording of high-resolution images of neuronal activity from the entire dorsal surface of the cortex in behaving animals (Makino et al. 2017, Chen et al. 2017, Allen et al. 2017), through the use of fluorescent calcium indicator molecules (Chen et al. 2013). Each pixel captures the fluorescence arising from dozens to thousands of neurons. Images can be steadily collected for hours at a high rate, from dozens to hundreds of Hertz. When the images are aggregated over time, a fluorescence trace can be extracted for each pixel. Inferring the time-variant neuronal activity from a given fluorescence trace is a challenging deconvolution problem.

Wide-field recording techniques date back a few decades. Before optogenetic manipulations were available, wide-field techniques used a combination of a single photon microscope and a camera to record the natural changes in illumination due to hemodynamic changes (Masino et al. 1993, for example). Because wide-field recordings cover large fields of view, the luminescence captured in each pixel originates from multiple sources (blood vessels and/or neurons). Hence the luminescence is strong enough to be captured through the skull, when the skull is relatively thin, as is the case for mice. This avoids invasive procedures, which in turn allows recordings of *in vivo* brain activity in behaving animals. It is also possible to conduct wide-field recordings in anesthetized animals (Kalchenko et al. 2014) or *in vitro* slices.

In recent years, the development of optogenetics has enabled neuronal recordings *in vivo* for long periods of time and across multiple days by combining optogenetic manipulations, which were originally developed for two photon microscopy, with chronically implanted windows that expose large fields of view. Furthermore, wide-field imaging can now be used to image neural activity across multiple brain regions (Silasi et al. 2016). As a result, wide-field imaging has rapidly become a standard recording technique (Clancy et al. 2019, Mann et al. 2017, Musall et al. 2018, Chen et al. 2017, Allen et al. 2017, Aimon et al. 2015, Wekselblatt et al. 2016, Makino et al. 2017). However, suitable approaches for deconvolution of wide-field calcium recordings are notably absent. In this paper, we develop a statistical model to extract mesoscale neural activity from wide-field recordings.

The problem of inferring the underlying neuronal activity from a fluorescence trace has recently been considered by a number of authors, in the case of fluorescence traces that result from the activity of a single neuron (Jewell & Witten 2018, Jewell et al. 2019, Friedrich et al. 2017, Pnevmatikakis et al. 2016). These papers make use of an auto-regressive model, originally proposed in Vogelstein et al. (2009), that associates the fluorescence y_t of a single neuron at the tth timepoint with the unobserved calcium c_t at the tth timepoint,

$$\begin{aligned} &y_t = \beta_0 + \beta_1 c_t + \epsilon_t, \qquad \epsilon_t \sim_{\text{ind.}} \left(0, \sigma^2\right), \quad t = 1, \ldots, T, \\ &c_t = \gamma c_{t-1} + s_t \quad t = 2, ..., T, \end{aligned} \tag{1}$$

where $s_t \geq 0$ allows for the occurrence of a spike at the tth timepoint, and where $\gamma \in (0, 1)$ is the rate of calcium decay. In words, (1) indicates that the calcium decays exponentially over time, unless there is a spike at the tth timepoint, in which case it increases; furthermore, the observed fluorescence is a noisy realization of the underlying calcium at each timepoint. In (1), β_0 corresponds to the baseline fluorescence, which must be estimated; however, we can set $\beta_1 = 1$ without loss of generality, as this simply amounts to scaling the calcium by a constant factor. To fit the model (1) with $\beta_1 = 1$, Friedrich et al. (2017), Jewell & Witten (2018), and Jewell et al. (2019) solve the optimization problem

$$\text{minimize}_{c_1,\ldots,c_T,\beta_0} \left\{ \sum_{t=1}^{T} (y_t - \beta_0 - c_t)^2 + \lambda P(c_t - \gamma c_{t-1}) \right\} \text{ subject to } c_t \geq \gamma c_{t-1},\ t = 2, \ldots, T, \tag{2}$$

where λ is a nonnegative tuning parameter, and where $P(\cdot)$ is a penalty function designed to induce sparsity in its argument — that is, to encourage $c_t - \gamma c_{t-1} = s_t = 0$ — so that at most timepoints, no spike is estimated to occur. Friedrich et al. (2017) makes use of an ℓ_1 penalty (Tibshirani 1996), while Jewell & Witten (2018) and Jewell et al. (2019) instead use an ℓ_0 penalty.

However, fluorescence traces that result from wide-field calcium imaging recordings correspond to the activity of a collection of neurons rather than the activity of a single neuron. Consequently, single-neuron deconvolution solutions, which assume that at most time points there are no spikes, cannot be directly applied. In this manuscript, we propose to extend the model (1), and the corresponding optimization problem (2), to the setting of wide-field calcium imaging recordings. We apply this new approach for deconvolution of wide-field calcium imaging recordings to data from the retrosplenial cortex (Swanson et al. 2018).

2 A Model for Wide-Field Recordings and Neuronal Activity

2.1 Extension of the Model (1) to Wide-Field Recordings

To begin, we consider the model (1), in a setting in which the observed fluorescence trace is the sum of the fluorescences associated with each of p neurons recorded at a given pixel. For the jth neuron, $j = 1, \ldots, p$, (1) takes the form

$$\begin{aligned} y_t^j &= \beta_0^j + c_t^j + \epsilon_t^j, \qquad \epsilon_t^j \sim_{\text{ind.}} \left(0, (\sigma^j)^2\right), \quad t = 1, \ldots, T, \\ c_t^j &= \gamma c_{t-1}^j + s_t^j, \quad t = 2, \ldots, T, \end{aligned} \tag{3}$$

where we assume that the rate of calcium decay, γ, is the same for all p neurons. However, in the wide-field case we do not separately observe the fluorescence for each of the p neurons; we instead observe their summed fluorescence. Summing (3) across the p neurons yields the model

$$\begin{aligned} \overline{y}_t &= \overline{\beta}_0 + \overline{c}_t + \overline{\epsilon}_t, \qquad \overline{\epsilon}_t \sim_{\text{ind.}} \left(0, \sum_{j=1}^p (\sigma^j)^2\right), \quad t = 1, \ldots, T, \\ \overline{c}_t &= \gamma \overline{c}_{t-1} + r_t, \quad t = 2, \ldots, T, \end{aligned} \tag{4}$$

where $\overline{y}_t \equiv \sum_{j=1}^p y_t^j$ is the total observed fluorescence of the p neurons, and where $\overline{\beta}_0$, $\overline{c}_t$, and $\overline{\epsilon}_t$ are the total baseline fluorescence, total calcium at the tth timepoint, and total noise at the tth timepoint, respectively. In (4), we can interpret r_t as the amount that the calcium increases, at the tth timepoint, as a result of spiking events in the p neurons; we will refer to this as the *spiking rate*. Because p is potentially quite large, on the order of hundreds to thousands of neurons, we do not expect r_t to be sparse. The total baseline fluorescence, $\overline{\beta}_0$, may result from the processing performed on the observed fluorescence; see e.g. the $\Delta F/F$ pre-processing of Chen et al. 2017.

2.2 Optimization Problem

The model (4) leads naturally to the optimization problem

$$\begin{aligned} &\text{minimize}_{\overline{c}_1, \ldots, \overline{c}_T, \overline{\beta}_0} \left\{ \sum_{t=1}^T (\overline{y}_t - \overline{\beta}_0 - \overline{c}_t)^2 + \lambda P(\overline{c}_2 - \gamma \overline{c}_1, \overline{c}_3 - \gamma \overline{c}_2, \ldots, \overline{c}_T - \gamma \overline{c}_{T-1}) \right\} \\ &\qquad \text{subject to } \overline{c}_t \geq \gamma \overline{c}_{t-1}, \ t = 2, \ldots, T. \end{aligned} \tag{5}$$

Problem (5) closely resembles the optimization problem (2) used to deconvolve the fluorescence trace for a single neuron (Friedrich et al. 2017, Jewell & Witten 2018, Jewell et al. 2019). For that task, the authors considered the use of a sparsity-inducing penalty, because a neuron is not expected to spike at most timepoints. By contrast, here we are considering a model in which $\overline{y}_t$ represents the total fluorescence at the tth timepoint summed over p neurons, and in which $\overline{c}_t - \gamma \overline{c}_{t-1}$ represents the spiking rate at the tth timepoint across all p neurons. Therefore, in the context of wide-field imaging, we do not want $P(\cdot)$ to be a sparsity-inducing penalty.

For convenience, in what follows, we will reparametrize (5) in terms of $r_1, \ldots, r_T$, where $r_t = \overline{c}_t - \gamma \overline{c}_{t-1}$ for $t = 2, \ldots, T$ as defined in (4), and where $r_1 \equiv \overline{c}_1$. The latter is strictly a definition intended for notational convenience. This reparametrizaton can be also expressed in matrix form,

$r = D\overline{c}$, where D is a $T \times T$ full-rank matrix with 1's on the diagonal and $-\gamma$'s just below the diagonal. This leads to rephrasing of the optimization problem (5) as follows:

$$\text{minimize}_{r_1,\ldots,r_T,\overline{\beta}_0} \left\{ \|\overline{y} - 1_T\overline{\beta}_0 - D^{-1}r\|^2 + \lambda P(r_2,\ldots,r_T) \right\} \text{ subject to } r_t \geq 0, t = 2,\ldots,T, \quad (6)$$

where 1_T is a vector of length T with all elements equal to 1.

Proposition 1. *The optimization problem* (5) *is equivalent to* (6)*, in the sense that* $\hat{\beta}_0 = \tilde{\beta}_0$ *and* $\hat{r} = D\tilde{c}$*, where* $\hat{r} = (\hat{r}_1,\ldots,\hat{r}_T)^\top$ *and* $\hat{\beta}_0$ *solve* (6)*, and* $\tilde{c} = (\tilde{c}_1,\ldots,\tilde{c}_T)^\top$ *and* $\tilde{\beta}_0$ *solve* (5)*.*

The proof is given in Appendix A.

We propose two possible forms for $P(\cdot)$:

1. *Dynamically-Binned Spiking Rate.* We consider the penalty

$$P(r_2,\ldots,r_T) = \sum_{t=3}^{T} |r_t - r_{t-1}|, \quad (7)$$

   which is a *fused lasso*, or *total variation denoising*, penalty (Tibshirani et al. 2005, Tibshirani et al. 2012, Condat 2013). This penalty encourages the spiking rate, r_t, to be constant over time, with only occasional changepoints, as shown in Figure 1B. This will yield estimates of the spiking rate that are constant within a bin, where the bins are themselves adaptively estimated from the data.

2. *Continuously-Varying Spiking Rate.* We consider the penalty

$$P(r_2,\ldots,r_T) = \sum_{t=3}^{T} (r_t - r_{t-1})^2, \quad (8)$$

   which encourages $r_t \approx r_{t-1}$, so that the spiking rate varies continuously over time, as shown in Figure 1C.

For simplicity, we can represent both penalties as

$$P_n(r_2,\ldots,r_T) = \sum_{t=3}^{T} |r_t - r_{t-1}|^n, \quad n = 1 \text{ or } 2, \quad (9)$$

where $n = 1$ corresponds to dynamically-binned (7) and $n = 2$ corresponds to a continuously-varying rate (8). The optimization problem (6) with penalty (9) can be written as

$$\text{minimize}_{r_1,\ldots,r_T,\overline{\beta}_0} \left\{ \|\overline{y} - 1_T\overline{\beta}_0 - D^{-1}r\|^2 + \lambda \sum_{t=3}^{T} |r_t - r_{t-1}|^n \right\} \text{ subject to } r_t \geq 0,\ t = 2,\ldots,T. \quad (10)$$

With $n = 1$ or $n = 2$, (10) is a convex optimization problem. However, it is not strictly convex and so the solution is not unique. As we prove in the following proposition, the solution to (10) is invariant under a constant shift in the spiking rate.

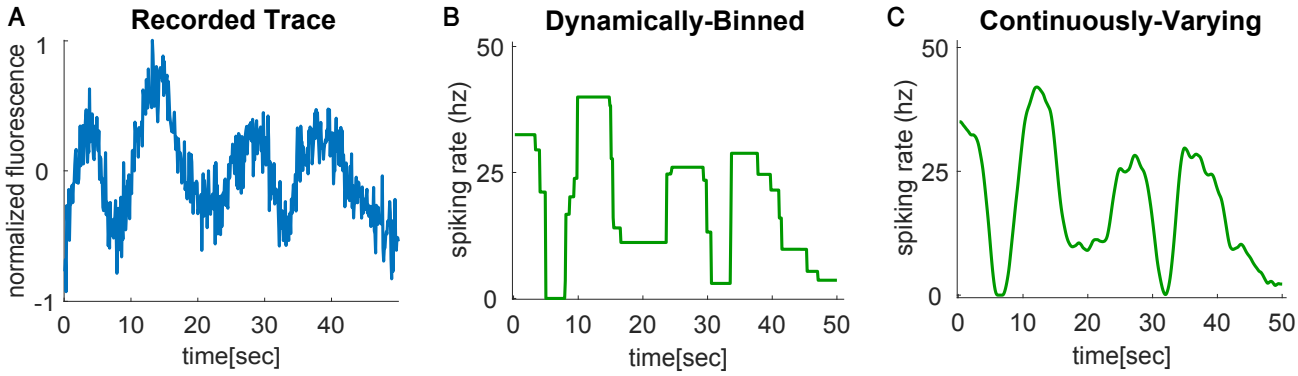

Figure 1: Illustrations of fluorescence trace and spiking rates estimated with two different penalties. A. An example of a recorded fluorescence trace. B. The spiking rate deconvolved from the recorded trace by solving (6) using the *Dynamically-Binned* penalty in (7). C. The spiking rate deconvolved from the recorded trace by solving (6) using the *Continuously-Varying* penalty in (8).

Proposition 2. *Let the pair* $(\hat{r}, \hat{\beta}_0)$ *denote a solution to problem* (10)*. Then* $\tilde{r}_t \equiv \hat{r}_t + d$ *for* $t = 2, \ldots, T$ *is also a solution to* (10) *for any* $d \geq -\min(r_t \; t = 2, \ldots, T)$*, provided that* $q(d, \hat{r}) \geq 0$*.*

We specify $q(d, r)$ as well as the values of $\tilde{r}_1$ and $\tilde{\beta}_0$ in the proof in Appendix C. The proof of Proposition 2 makes use of Proposition 3.

Proposition 3. *The pair* $(\tilde{r}, \tilde{\beta}_0)$ *is a solution to problem* (10) *if and only if* $\tilde{\beta}_0 = \frac{1}{T} 1_T^\top (\overline{y} - D^{-1}\tilde{r})$ *and* $\tilde{r}$ *is a solution to*

$$\text{minimize}_{r_1,\ldots,r_T} \left\{ \|\tilde{y} - Ar\|^2 + \lambda \sum_{t=3}^{T} |r_t - r_{t-1}|^n \right\} \text{ subject to } r_t \geq 0, \; t = 2, \ldots, T, \tag{11}$$

where $A = \left(I - \frac{1}{T} 1_T 1_T^\top\right) D^{-1}$ *and* $\tilde{y} = \left(I - \frac{1}{T} 1_T 1_T^\top\right) \overline{y}$*, with* I *the* $T \times T$ *identity matrix and* $1_T 1_T^\top$ *a* $T \times T$ *matrix of ones.*

Note that $\tilde{y}$ is the mean subtracted version of $\overline{y}$, and A is the column-wise mean subtracted version of D^{-1}.

We prove Proposition 3 in Appendix B. Proposition 3 implies that we can solve (10) by simply solving (11) for r and then using a closed form expression for $\overline{\beta}_0$.

3 Algorithm

3.1 Overview of Proximal Gradient Descent

Equation (11) with $n \geq 1$ is a convex optimization problem (Boyd & Vandenberghe 2004) for which a number of efficient algorithms are available to solve for the global optimum. Here, we make use of a proximal gradient descent algorithm, which provides an efficient and flexible framework for solving this problem.

We now provide a brief overview of proximal gradient descent; a detailed treatment can be found in Parikh et al. (2014). Suppose that we wish to solve the optimization problem

$$\underset{x}{\text{minimize}} \, \{f(x) + g(x)\} \tag{12}$$

where $f(\cdot)$ is a smooth convex function, and $g(\cdot)$ is convex but possibly non-differentiable. Then, under mild conditions, an iterative algorithm that initializes x at $x^{(0)}$, and then at the tth iteration applies the update

$$x^{(t)} \leftarrow \text{Prox}_{sg(\cdot)}\left(x^{(t-1)} - s\nabla f(x^{(t-1)})\right),$$

will converge to the global optimum. Here, s is a stepsize chosen so that $s \leq 1/L$, where L is the Lipschitz constant for the function $\nabla f(\cdot)$. The notation $\text{Prox}_{sg(\cdot)}$ indicates the *proximal operator* of the function $sg(\cdot)$, defined as

$$\text{Prox}_{sg(\cdot)}(\tilde{x}) \equiv \arg\min_x \left\{\frac{1}{2}\|x - \tilde{x}\|^2 + sg(x)\right\}. \tag{13}$$

Therefore, proximal gradient descent provides a simple recipe for solving a broad class of convex optimization problems of the form (12), provided that the proximal operator (13) is easily computed, and the function $\nabla f(\cdot)$ is Lipschitz continuous.

3.2 Algorithm for Dynamically-Binned Spiking Rate

We now propose a proximal gradient descent algorithm for solving (10) with $n = 1$. By Proposition 3, we can simply solve

$$\text{minimize}_{r_1,\ldots,r_T} \left\{\|\tilde{y} - Ar\|^2 + \lambda \sum_{t=3}^{T} |r_t - r_{t-1}|\right\} \text{ subject to } r_t \geq 0,\ t = 2, \ldots, T. \tag{14}$$

Using the notation of (12), we take $f(r) = \|\tilde{y} - Ar\|^2$ and $g(r) = \lambda \sum_{t=3}^{T} |r_t - r_{t-1}|^n$.

The following results, which are proven in Appendices D and E, will be useful.

Proposition 4. *The function $\nabla_r f(r) = -2A^\top(\tilde{y} - Ar)$ is Lipschitz continuous with Lipschitz constant L satisfying $L \leq 2\left(\frac{1-\gamma^T}{1-\gamma}\right)^2$.*

Proposition 5. *Let $\hat{r}$ solve the optimization problem*

$$\text{minimize}_{r_1,\ldots,r_T} \left\{\frac{1}{2}\|r - \tilde{x}\|^2 + s\lambda \sum_{t=3}^{T} |r_t - r_{t-1}|\right\}. \tag{15}$$

Then $\tilde{r}_t \equiv \max(\hat{r}_t, 0)$ for $t = 2, \ldots, T$ and $\tilde{r}_1 \equiv \hat{r}_1$ solves the optimization problem

$$\text{minimize}_{r_1,\ldots,r_T} \left\{\frac{1}{2}\|r - \tilde{x}\|^2 + s\lambda \sum_{t=3}^{T} |r_t - r_{t-1}|\right\} \textit{ subject to } r_t \geq 0,\ t = 2, \ldots, T. \tag{16}$$

A number of standard solvers, such as the `flsa` solver in `R` (Hoefling 2010), are available to solve (15). We solve (15) by implementing the proposal of Condat (2013). Proposition 5 implies that given a solution to (15), solving (16) is straightforward.

Propositions 3–5 lead directly to Algorithm 1 for solving (14).

Algorithm 1 Dynamically-Binned Rate Deconvolution: Solving (10) with $n = 1$

Given the signal $\overline{y}$ and the calcium decay in a time step γ, do:

1. Calculate $\tilde{y} = \left(I - \frac{1}{T}1_T 1_T^\top\right)\overline{y}$.
2. Set step size $s = \frac{1}{2}\left(\frac{1-\gamma}{1-\gamma^T}\right)^2$, and $A = \left(I - \frac{1}{T}1_T 1_T^\top\right)D^{-1}$.
3. Initialize $r \in \mathbb{R}_+^T$.
4. Iterate until convergence:
   (a) Let $\tilde{x} \leftarrow r + 2sA^\top(\tilde{y} - Ar)$.
   (b) Let $r \leftarrow \text{argmin}_r \left\{\frac{1}{2}\|\tilde{x} - r\|^2 + s\lambda \sum_{t=3}^T |r_t - r_{t-1}|\right\}$.
   (c) Let $r_t \leftarrow (r_t)_+$ for $t = 2, \ldots, T$.
5. Let $\overline{\beta}_0 = \frac{1}{T}(1_T)^\top(\overline{y} - D^{-1}r)$.

3.3 Algorithm for Continuously-Varying Spiking Rate

We now propose a proximal gradient descent algorithm for solving (10) with $n = 2$. By Proposition 3 it suffices to solve

$$\text{minimize}_{r_1,\ldots,r_T} \left\{\|\tilde{y} - Ar\|^2 + \lambda \sum_{t=3}^T (r_t - r_{t-1})^2\right\} \text{ subject to } r_t \geq 0,\ t = 2, \ldots, T. \tag{17}$$

We can further express (17) as

$$\text{minimize}_{r_1,\ldots,r_T} \left\{\|\tilde{y} - Ar\|^2 + \lambda \sum_{t=3}^T (r_t - r_{t-1})^2 + \sum_{t=2}^T h(r_t)\right\}. \tag{18}$$

with

$$h(r) = \begin{cases} 0, & \text{if } r \geq 0 \\ \infty, & \text{if } r < 0. \end{cases} \tag{19}$$

This allows us to express the objective function in (17) as $f(r) + g(r)$ where

$$f(r) = \|\tilde{y} - Ar\|^2 + \sum_{t=3}^T (r_t - r_{t-1})^2 \text{ and } g(r) = \sum_{t=2}^T h(r_t). \tag{20}$$

Proposition 6. *The function $\nabla f(r) = -2A^\top(\tilde{y} - Ar) + 2\lambda Z r$ is Lipschitz continuous with Lipschitz constant L satisfying $L \leq 2\frac{\left(1-\gamma^T\right)^2 + 4\lambda(1-\gamma)^2}{(1-\gamma)^2}$, where*

$$Z_{i,j} = \begin{cases} -1, & \textit{if } (j = i-1 \textit{ or } j = i+1) \textit{ and } (i + j > 3) \\ 2, & \textit{if } i = j \textit{ and } 2 < i < T \\ 1, & \textit{if } i = j \textit{ and } (i = 2 \textit{ or } i = T) \\ 0, & \textit{otherwise} \end{cases}. \tag{21}$$

We note that $\sum_{t=3}^{T}(r_t - r_{t-1})^2 = r^\top Z r$. Proposition 6 is proven in Appendix F.

Proposition 7. *For any $s \geq 0$, the solution to the optimization problem*

$$\text{minimize}_{r_1,\ldots,r_T} \left\{ \frac{1}{2}\|r - \tilde{x}\|^2 + s\sum_{t=2}^{T} h(r) \right\} \tag{22}$$

is given by $r_t = \max(\tilde{x}_t, 0)$ for $t = 2, \ldots, T$ and $r_1 = \tilde{x}_1$.

Propositions 3, 6 and 7 lead to Algorithm 2 for solving (17).

Algorithm 2 Continuously-Varying Rate Deconvolution: Solving (10) with $n = 2$

Given the signal $\overline{y}$ and the calcium decay in a time step γ, do:

1. Calculate $\tilde{y} = \left(I - \frac{1}{T}1_T 1_T^\top\right)\overline{y}$.
2. Set step size $s = \frac{1}{2}\frac{(1-\gamma)^2}{(1-\gamma^T)^2 + 4\lambda(1-\gamma)^2}$, and $A = \left(I - \frac{1}{T}1_T 1_T^\top\right) D^{-1}$.
3. Initialize $r \in \mathbb{R}_+^T$.
4. Iterate until convergence:
   (a) Let $r \leftarrow r + 2sA^\top(\tilde{y} - Ar) - 2s\lambda Z r$.
   (b) Let $r_t \leftarrow (r_t)_+$ for $t = 2, \ldots, T$.
5. Let $\overline{\beta}_0 = \frac{1}{T}(1_T)^\top(\overline{y} - D^{-1}r)$.

In Appendix G we consider solving (10) with $n = 2$ in the absence of the non-negativity constraint on r.

4 Results on Simulated Data

4.1 Evaluation of Algorithms 1 and 2

Here we evaluate the performances of Algorithms 1 and 2 on simulated fluoresence data generated from randomly drawn rate traces. We measure the difference between the true rate r and the deconvolved rate r^{est} as

$$\text{err}(r, r^{est}) = \frac{1}{T-1}\sum_{t=2}^{T} |r_t - r_t^{est}|. \tag{23}$$

Since the solution to (10) is unique only up to a contant shift (see Proposition 2), we compute (23) after subtracting the mean from both r_t and r_t^{est} for $t = 2, \ldots, T$.

We begin by generating 500 random piece-wise constant rate traces and 500 random continuous rate traces, and used those to calculate calcium and fluorescence traces (details are given in Appendix H; see Figures 2A, 2B, 2D, 2E). We used Algorithms 1 and 2 to deconvolve the fluorescence traces generated from piece-wise constant and continuous rates, respectively.

For each fluorescence trace, we performed the deconvolution for a range of values of the penalty λ and the decay rate γ. For each pair of λ and γ, we calculated the average error (23) across all 500 simulated data sets, as well as the average of $\sum_{t=3}^{T} |r_t - r_{t-1}|^n$. The results are summarized in Figure 2C and 2E.

Our results indicate that deconvolving the rate using the original calcium decay constant γ indeed yields the lowest error for most values of the penalty λ.

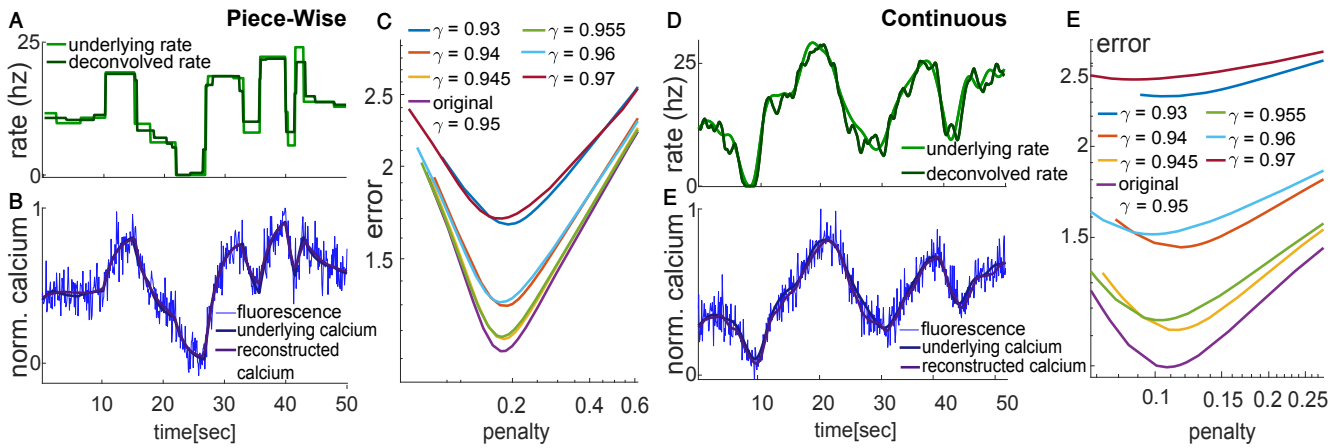

Figure 2: Deconvolution of the fluorescence signal into spiking rate. A–C. Dynamically-binned deconvolution. A,B. Underlying piece-wise constant rate (light green) was created by the random process described in Appendix H. The rate was transformed, using the model (4), into the underlying calcium trace (dark blue). White noise and a constant shift were added to the latter to create the noisy fluorescence signal (blue, plotted here without the shift). Algorithm 1 was used to estimate the spiking rate. The resulting estimated rate (dark green) successfully recovers the underlying rate. The estimated rate was used to estimate the underlying calcium trace (purple) using (4). C. Error of the deconvolution algorithm, using (23) and a range of values of γ and λ, averaged over 500 simulated data sets. D–F. Continuously-varying rate deconvolution. Algorithm 2 was used to estimate the spiking rate; details are as in A–C.

4.2 Comparisons with Other Methods

We compare our algorithms' performances to those of other approaches that can be used to analyze wide-field calcium imaging data.

1. *First Differences Deconvolution.* This method assumes that the calcium is represented directly by the fluorescence. Hence, the rate can be found by assuming $\overline{c} = \overline{y}$ in (4), yielding

$$r_t = \overline{y}_t - \gamma \overline{y}_{t-1} \text{ for } t = 2, \ldots, T. \tag{24}$$

   In the special case of $\gamma = 1$, this method is equivalent to estimating the rate by computing the first differences of the observed fluorescence. We used a moving average window to smooth the resulting estimated rate, and added a mean shift to the result in order to avoid negative rates. This method is a natural and simple alternative to our approach: it bypasses estimation of the underlying calcium, and replaces the penalties in (7) and (8) with a simple *post hoc* smoothing.

2. *Lucy-Richardson Deconvolution.* This iterative algorithm was originally used in astrophysics to restore the light source from a filtered blurred image (Lucy 1974, Richardson 1972). It assumes that the blurring is a result of a convolution by a filter f and additive noise. The algorithm maximizes the likelihood of the original image and the cleared image, assuming the noise has a Poisson distribution. In wide-field calcium imaging, this approach has been applied with a one-dimensional filter given by $f_t = \gamma^t$ (Wekselblatt et al. 2016). To improve performance, we smooth the results using a moving average. The algorithm guarantees non-negativity of the estimated rate.

We compare the above methods and Algorithms 1 and 2 by simulating continuously-varying spiking rates (Appendix H) and measuring the errors (23) of the deconvolved rates. Results are shown in Figure 3 for a range of values of γ and λ, averaged over 500 simulated data sets. We also measured the fluctuation in each of the deconvolved rate traces by

$$\zeta(r) = \sum_{t=3}^{T} (r_t - r_{t-1})^2, \tag{25}$$

which is the penalty used in (10) with $n = 2$.

We find that both the first differences and Lucy-Richardson approaches perform best when the true value of γ is used (Figures 3B and 3D). However, the best results for the first differences approach are achieved with larger fluctuations compared to other methods (Figures 3B and 3E). This is also evident from inspecting the best deconvolved rate for the example trace (Figure 3A). Lucy-Richardson achieves the best fit with fewer fluctuations (Figure 3D), which results in a smoother deconvolved rate (Figure 3C).

Figures 3E and 3F indicate that our continuously-varying spiking rate algorithm performs the best, with the smallest error and the smallest fluctuations at its minimum error (with $\gamma = 0.95$ and $\lambda = 2750$). Lucy-Richardson comes in second, just ahead of our dynamically-binned algorithm, and far ahead of the first differences method. When we simulated data with a piecewise constant spiking rate (not shown), our dynamically-binned algorithm performed the best, with Lucy-Richardson second and just ahead of our continuously-varying algorithm, and the first differences method once again far behind.

Last, we calculated the correlations of the deconvolved rate traces, the underlying calcium, and the fluorescence with the underlying rate traces. We see that all deconvolution methods yield substantially better estimates of the underlying rate than simply using the underlying calcium or observed fluorescence.

5 Results on Recorded Data

5.1 Whole Dorsal Surface Recordings

Here we evaluate the performances of Algorithms 1 and 2 and the first differences and Lucy-Richardson deconvolution methods on a dataset described in Musall et al. (2018). The dataset consists of fluorescence traces recorded simultaneously at 20Hz from the whole dorsal surface of a genetically encoded GCaMP6s mouse. During the one hour recording, the mouse performed multiple trials of some task. Dimension reduction (Musall et al. 2018), hemodynamics removal (Wekselblatt et al. 2016), and $\Delta F/F$ transformation were performed on the traces. Qualitatively

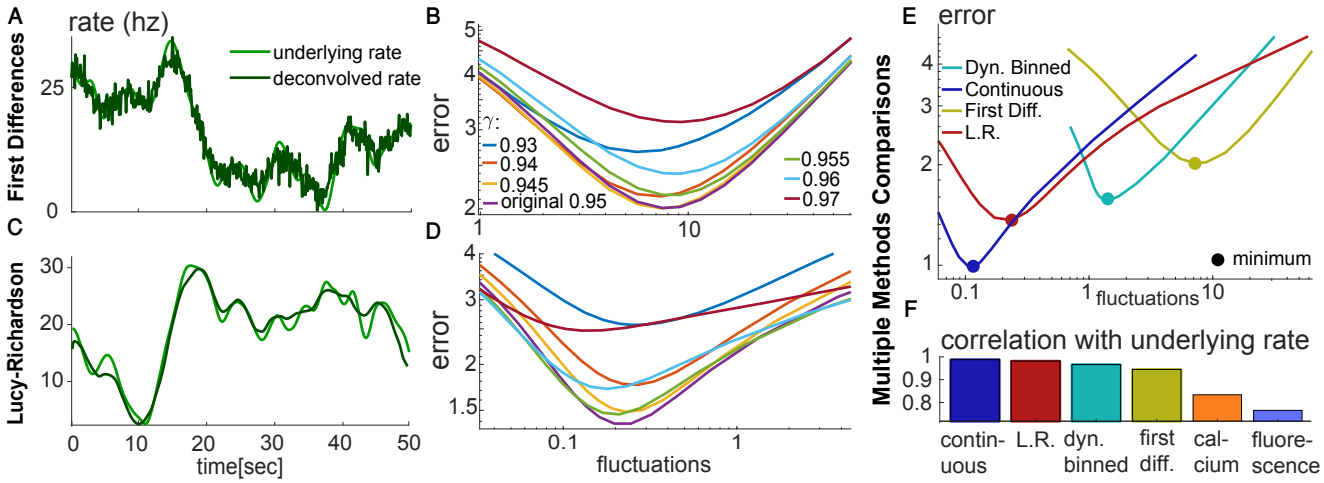

Figure 3: Deconvolution using competing approaches. A–B. Deconvolution of continuously-varying spiking rate using the first differences method. A. An example: Underlying continuous rate (light green) and estimated rate using the first differences method (dark green). The example was generated using the parameter set that yields the lowest average error for this method. B. Error of the first differences method quantified using (23). The x-axis displays the fluctuations in the estimated rate defined in (25). We used a range of γ values and a variety of smoothing window lengths. C–D. Similar to A–B but for the Lucy-Richardson (L.R.) method. E. Comparisons of Algorithms 1 and 2 with other methods. F. Correlation of the underlying rate with the deconvolved rate achieved by our algorithms, other methods, or the underlying calcium or fluorescence.

similar results are achieved after performing only a $\Delta F/F$ transformation prior to deconvolution (not shown).

In this data set, because the true spiking rate is unknown, we split each recorded fluorescence trace $\overline{y} = (\overline{y}_1, \overline{y}_2, \ldots, \overline{y}_T)$ into "odd" and "even" traces, $\overline{y}^{odd} = \left(\overline{y}_1, \overline{y}_3, \ldots, \overline{y}_{T-1}\right)$ and $\overline{y}^{even} = (\overline{y}_2, \overline{y}_4, \ldots, \overline{y}_T)$. We then deconvolve the odd trace $\overline{y}^{odd}$ to estimate the spiking rate r^{odd} and the intercept, $\overline{\beta}_0^{odd} = \frac{2}{T} 1_{\frac{T}{2}}^{\top} (\overline{y}^{odd} - D^{-1} r^{odd})$. To calculate D, defined in Section 2.2, we use $\gamma = 0.95$, the estimated decay (Chen et al. 2013) for GCaMP6s mice at the 10Hz frequency corresponding to the "odd" and "even" traces.

Next, we apply the second line in (4), adapted to the odd observations, in order to compute $\overline{c}^{odd}$ according to $\overline{c}_t^{odd} = \gamma \overline{c}_{t-2}^{odd} + r^{odd}$ for $t = 1, 3, 5, \ldots, t-1$. We then apply the first line in (4) to compute

$$\mathrm{err}(\overline{y}^{even}, \overline{c}^{odd}) = \frac{2}{T\left(\max(\overline{y}) - \min(\overline{y})\right)} \sum_{t=1}^{T/2} |\overline{y}_{2t}^{even} - \overline{\beta}_0^{odd} - \overline{c}_{2t-1}^{odd}|. \tag{26}$$

The four different deconvolution methods were applied for a range of tuning parameter values. We find that Algorithm 2 results in the smallest error out of all methods; see Figure 4J. The first differences method as well as Algorithm 1 result in very similar performances, while the latter has less fluctuation in the deconvolved rate; see Figure 4E-J.

5.2 Parallel Wide-Field and Spike Recordings

Here we compare the performances of Algorithm 1, Algorithm 2, first differences and Lucy-Richardson on parallel recorded spiking data described in Clancy et al. (2019). The data consists

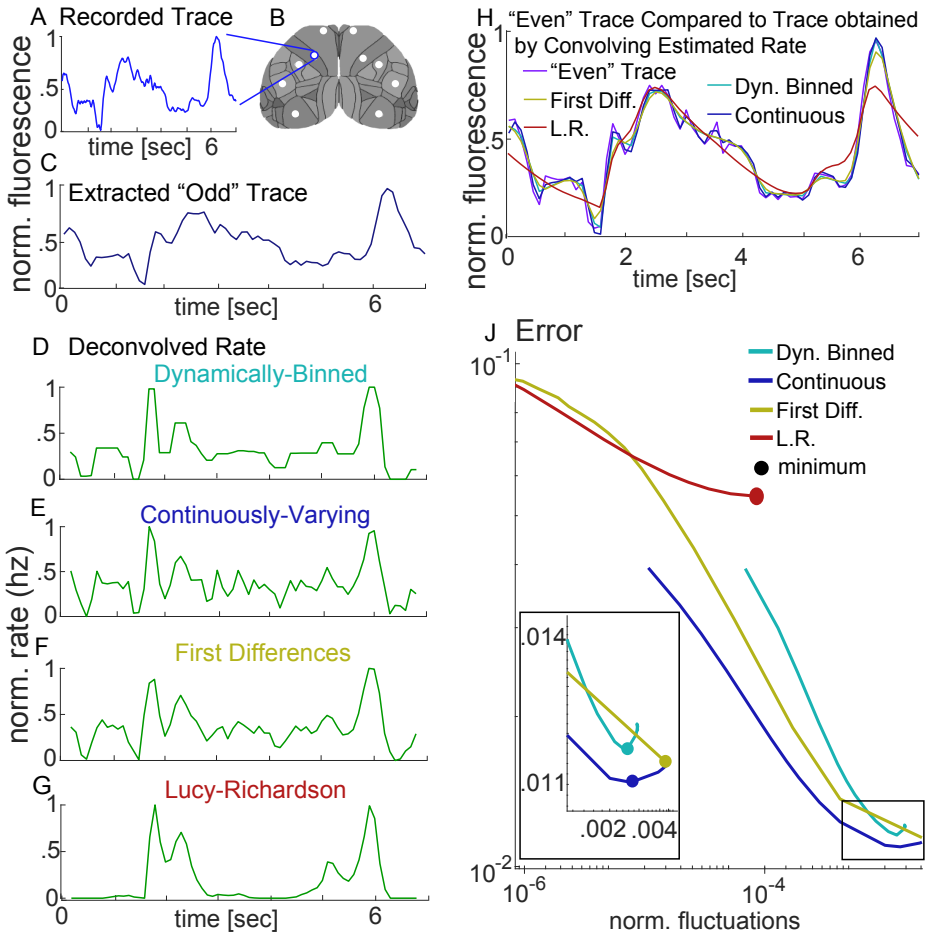

Figure 4: Deconvolution of recorded data from Musall et al. (2018). A. An example of a fluorescence trace recorded from a pixel during a single trial. B. Schematic of a whole dorsal surface wide-field view. White circles indicate the locations of pixels from which fluorescence traces were recorded. The example in A was taken from the pixel circled in blue. C. A fluorescence trace created by extracting from A. the odd time points. D-G. Deconvolution of the fluorescence trace in C. into spiking rate using the D. dynamically-binned algorithm (Algorithm 1), E. continuously-varying algorithm (Algorithm 2), F. first differences method, and G. Lucy-Richardson method. The rates in D-G are displayed with tuning parameters selected to minimize the error in J. H. The "even" trace from A is displayed alongside the results of convolving the spiking rates in D-G obtained from the "odd" trace. J. The relative error between the "even" fluorescence trace and the convolved calcium from the deconvolved spiking rate using the "odd" fluorescence trace (26). The error is displayed as a function of the normalized fluctuations, defined by (25) divided by $\max(r^{odd})^2$. Different fluctuation levels and errors were achieved by varying λ for the dynamically-binned and continuously-varying algorithms, and by varying the smoothing window for the first differences and Lucy-Richardson methods. Larger fluctuation values correspond to lower values of λ and smaller smoothing windows. When $\lambda = 0$,then no smoothing occurs; this corresponds to the right hand side of the figure. Results were averaged across traces from all 10 pixels highlighted in B. and 100 trials for each pixel. All fluorescence traces were shifted to have only positive values prior to performing the Lucy-Richardson method.

of the number of spikes recorded from a multi-channel silicon probe in V1 in each 25ms bin (which we interpret as giving the true spiking rate) and, simultaneously, a fluorescence trace recorded from the same location at 40Hz. A $\Delta F/F$ transformation was applied to the fluorescence trace after recording. We note that the probe detects only a subset of the spikes that contribute to the calcium fluorescence. To test the different deconvolution methods, we use the rate of the detected spikes as an estimate of the underlying spiking rate.

To evaluate the methods, we calculated the difference between the deconvolved and the recorded spike rates (23), after dividing each spike rate by its standard deviation.

A decay rate of $\gamma = 0.97$ has been reported in the literature (Clancy et al. 2019, Chen et al. 2013) for GCaMP6f mice recorded at 40Hz. We investigated a range of values of γ and found that all methods attained the lowest error (23) with $\gamma = 0.975$ except Lucy-Richardson, which attained the lowest error with $\gamma = 0.98$. We also found that the Lucy-Richardson method attained its lowest error when its fluctuation (defined in (25)) is quite small, Figure 5D. This agrees with results shown in Figure 4G,J. The smallest error, across all methods and parameter values, was achieved by the dynamically-binned algorithm, Figure 5D. Comparable lowest error values were achieved by the continuously-varying and Lucy-Richardson methods. Moreover, the dynamically-binned and continuously-varying algorithms achieved these values with greater fluctuations compared to Lucy-Richardson, suggesting that they are more sensitive to the original calcium fluctuations.

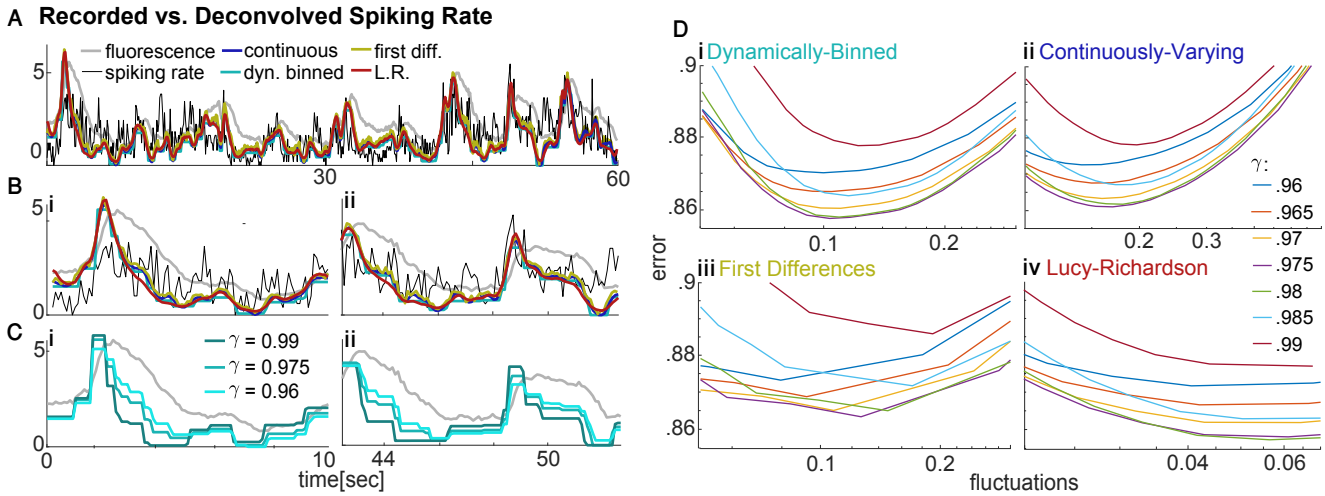

Figure 5: Deconvolved spiking rate compared with recorded spiking rate. A. An example of spiking rate (black) recorded in parallel to the fluorescence (gray), along with the rates estimated by the four deconvolution methods. Tuning parameters were selected to yield the lowest error (23). B. Enlarged regions from A for (i) 0-10 sec and (ii) 42-52 sec. C. Deconvolved spiking rate (shades of turquoise) calculated using different calcium decay values, γ, to deconvolve the fluorescence trace (gray) for (i) 0-10 sec, and (ii) 42-52 sec, used for the dynamically-binned algorithm. D. The error (23) between the recorded spiking rate and the estimated spiking rate obtained using (i) the dynamically-binned algorithm, (ii) the continuously-varying algorithm, (iii) first differences, and (iv) Lucy-Richardson. For each of the methods, the error was calculated for a range of tuning parameter values. Larger fluctuation values, defined in (25), correspond to lower values of λ for Algorithms 1 and 2 and smaller smoothing windows for first differences and Lucy-Richardson methods. Results were averaged across 75 traces, each of which was 60 seconds long.

6 Code Availability

To support further research we make our code available at https://github.com/meravstr/Wide-Field-Deconvolution.

7 Acknowledgements

We are grateful for Kelly Clancy who supplied us with parallel fluorescence spiking recordings taken at Thomas Mrsic-Flogel lab. We also thank Simon Musall from Anne Churchland's lab for supplying fluorescence traces. MS thanks Nili Guttmann-Beck, Amir Beck Israel Nelken and Matteo Carandini for helpful notes and OCNC for hosting during the initiation of this project. DMW has been supported by NSF CAREER DMS-1252624, Simons Investigator Award in Mathematical Modeling of Living Systems, NIH grant DP5OD009145, NIH R01 EB026908 and NIH R01 DA047869. We also acknowledge support from the Sackler Foundation, the Swartz Foundation via the Swartz Center for Theoretical Neuroscience at the University of Washington, and the NSF-ERC Center for Sensorimotor Neural Engineering at the University of Washington and NSF-DMS Grants 1514743 and CAREER-1056125. We thank the Allen Institute founders, Paul G. Allen and Jody Allen, for their vision, encouragement and support.

A Proof of Proposition 1

Proof. The result follows by noticing that for the one-to-one mapping $r = D\overline{c}$, the optimization problems (5) and (6) are identical. □

B Proof of Proposition 3

Lemma 1. *Let* $A = PD^{-1}$, $\tilde{y} = P\overline{y}$ *with* $P = I - \frac{1}{T}1_T 1_T^\top$ *and* $\overline{\beta}_0 = \frac{1}{T}\sum_{t=1}^{T}(\overline{y}_t - (D^{-1}r)_t)$. *Then,* $\|\overline{y} - 1_T\overline{\beta}_0 - D^{-1}r\|^2 = \|\tilde{y} - Ar\|^2$.

Proof of Lemma 1. Since $P^\top P = P$ and $P1_T = 0$, it follows that

$$\|\overline{y} - 1_T\overline{\beta}_0 - D^{-1}r\|^2 = \left(\overline{y} - 1_T\overline{\beta}_0 - D^{-1}r\right)^\top \left(I - \frac{1}{T}1_T 1_T^\top + \frac{1}{T}1_T 1_T^\top\right)\left(\overline{y} - 1_T\overline{\beta}_0 - D^{-1}r\right)$$

$$= \left(\overline{y} - 1_T\overline{\beta}_0 - D^{-1}r\right)^\top P^\top P\left(\overline{y} - 1_T\overline{\beta}_0 - D^{-1}r\right) + \left(\overline{y} - 1_T\overline{\beta}_0 - D^{-1}r\right)^T \frac{1}{T}1_T 1_T^\top\left(\overline{y} - 1_T\overline{\beta}_0 - D^{-1}r\right)$$

$$= \|\tilde{y} - Ar\|^2 + \frac{1}{T}\|1_T^\top\overline{y} - 1_T^\top 1_T\overline{\beta}_0 - 1_T^\top D^{-1}r\|^2 = \|\tilde{y} - Ar\|^2.$$

The last equality follows from the fact that $1_T^\top\overline{y} - 1_T^\top 1_T\overline{\beta}_0 - 1_T^\top D^{-1}r = 0$. □

Proof of Proposition 3. Taking the derivative of (10) with respect to $\overline{\beta}_0$ and setting it equal to zero, we find that $\overline{\beta}_0 = \frac{1}{T}\sum_{t=1}^{T}(\overline{y}_t - (D^{-1}r)_t)$. The result follows from Lemma 1. □

C Proof of Proposition 2

Proof of Proposition 2. It follows from Proposition 3 that $\hat{r}$ is a solution to (10) if and only if it is also a solution to (11). Hence, $\hat{r}$ minimizes the objective function $f(r) = \|\tilde{y} - Ar\|^2 + \lambda \sum_{t=3}^{T} \|r_t - r_{t-1}\|^n$. Therefore, any $\tilde{r}$ that satisfies $f(\tilde{r}) = f(\hat{r})$ with $\tilde{r}_t > 0$ for $t = 2, \ldots, T$ is a solution to the optimization problem (11) as well.

If we construct $\tilde{r}_t = \hat{r}_t + d$ for $t = 2, \ldots, T$ with $d \geq -\min(\hat{r}_t \; t = 2, \ldots, T)$, then $\tilde{r}_t \geq 0$ for $t = 2, \ldots, T$. In addition, $\lambda \sum_{t=3}^{T} |\tilde{r}_t - \tilde{r}_{t-1}| = \lambda \sum_{t=3}^{T} |\hat{r}_t - \hat{r}_{t-1}|$, so to show that $f(\tilde{r}) = f(\hat{r})$ it suffices to show that $\|\tilde{y} - A\tilde{r}\|^2 = \|\tilde{y} - A\hat{r}\|^2$. We will do so by choosing an appropriate value for $\tilde{r}_1$. In particular, notice that

$$\begin{aligned} \|\tilde{y} - A\tilde{r}\|^2 &= \sum_{t=1}^{T} \left(\tilde{y}_t - \sum_{j=1}^{T} (A_{t,j}\tilde{r}_j) \right)^2 = \sum_{t=1}^{T} \left(\tilde{y}_t - A_{t,1}\tilde{r}_1 - \sum_{j=2}^{T} [A_{t,j}(\hat{r}_j + d)] \right)^2 \\ &= \tilde{r}_1^2 \sum_{t=1}^{T} A_{t,1}^2 - 2\tilde{r}_1 \sum_{t=1}^{T} A_{t,1} \left(\tilde{y}_t - \sum_{j=2}^{T} [A_{t,j}(\hat{r}_j + d)] \right) + \sum_{t=1}^{T} \left(\tilde{y}_t - \sum_{j=2}^{T} [A_{t,j}(\hat{r}_j + d)] \right)^2 . \end{aligned} \tag{27}$$

So it suffices to find the value of $\tilde{r}_1$ for which (27) equals $\|\tilde{y} - A\hat{r}\|^2$. We observe that (27) is quadratic in $\tilde{r}_1$. We define

$$\begin{aligned} a &\equiv \sum_{t=1}^{T} A_{t,1}^2, \\ b &\equiv -2 \sum_{t=1}^{T} A_{t,1} \left(\tilde{y}_t - \sum_{j=2}^{T} [A_{t,j}(\hat{r}_j + d)] \right), \\ c &\equiv \sum_{t=1}^{T} \left(\tilde{y}_t - \sum_{j=2}^{T} [A_{t,j}(\hat{r}_j + d)] \right)^2 - \|\tilde{y} - A\hat{r}\|^2. \end{aligned}$$

A solution exists provided that $b^2 - 4ac \geq 0$, in which case $\tilde{r}_1 = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$.

From Proposition 3, we know that if $\tilde{r}$ is a solution to (11), then the pair $(\tilde{r}, \tilde{\beta}_0)$ is a solution to (10) with $\tilde{\beta}_0 = \frac{1}{T} 1_T^\top (\overline{y} - D^{-1}\tilde{r})$. Defining $q(d, r) \equiv b^2 - 4ac$, the proof is complete.

□

D Proof of Proposition 4

Equation (4) implies that

$$\overline{c}_t = \sum_{t'=1}^{t} \gamma^{t-t'} r_{t'}. \tag{28}$$

Since $\overline{c} = D^{-1}r$, we find from (28) that the entries of D^{-1} are given by

$$\left(D^{-1}\right)_{i,j} = \begin{cases} 0, & \text{if } i < j \\ 1, & \text{if } i = j \\ \gamma^{i-j}, & \text{if } i > j \end{cases}. \tag{29}$$

Lemma 2. $\max_{i=1,\dots,T}\left\{\sum_{j=1}^{T}\left(D^{-1}\right)_{i,j}\right\} \leq \frac{1-\gamma^T}{1-\gamma}$.

Proof. Using (29) for $i = 1, \dots, T$:

$$\sum_{j=1}^{T}\left(D^{-1}\right)_{i,j} = \sum_{j=1}^{i}\gamma^{i-j} = \sum_{j=T-i+1}^{T}\gamma^{T-j} \leq \sum_{j=1}^{T}\gamma^{T-j} = \frac{1-\gamma^T}{1-\gamma},$$

where the last step was calculated using the expression for the sum of a geometric series. □

Following the same reasoning, one can also show that $\max_{j=1\dots T}\left\{\sum_{i=1}^{T}\left(D^{-1}\right)_{i,j}\right\} \leq \frac{1-\gamma^T}{1-\gamma}$.

Proof of Proposition 4. The Lipschitz constant of $-2A^\top(\tilde{y} - Ar)$ is given by the largest eigenvalue of $2A^\top A$ (Bubeck et al. 2015). To find it, it is convenient to first explore the largest eigenvalue of $\left(D^{-1}\right)^\top D^{-1}$, which is a symmetric matrix with positive real entries. By the Perron Frobenius Theorem and Lemma 2, we can bound its largest eigenvalue by its largest single row sum:

$$\begin{aligned}\Lambda_{\max}\left(\left(D^{-1}\right)^\top D^{-1}\right) &\leq \max_i \sum_{j=1}^{T}\sum_{k=1}^{T}\left(D^{-1}\right)_{k,i}\left(D^{-1}\right)_{k,j} = \max_i \sum_{k=1}^{T}\left(D^{-1}\right)_{k,i}\left(\sum_{j=1}^{T}\left(D^{-1}\right)_{k,j}\right)\\ &\leq \max_i \sum_{k=1}^{T}\left(D^{-1}\right)_{k,i}\left(\frac{1-\gamma^T}{1-\gamma}\right) \leq \left(\frac{1-\gamma^T}{1-\gamma}\right)^2.\end{aligned} \tag{30}$$

We recall that $A = PD^{-1}$ and we observe that $P^\top P = P$, where $P = \left(I - \frac{1}{T}1_T 1_T^\top\right)$, with I the identity matrix and $1_T 1_T^\top$ a $T \times T$ matrix of ones. Together with (30) and Rayleigh quotient properties, the following holds for any vector u:

$$\begin{aligned}\frac{u^\top 2A^\top A u}{u^\top u} &= 2\frac{u^\top\left(D^{-1}\right)^\top P D^{-1} u}{u^\top u} = 2\left[\frac{u^\top\left(D^{-1}\right)^\top D^{-1}u}{u^\top u} - \frac{1}{T}\frac{u^\top 1_T 1_T^\top u}{u^\top u}\right]\\ &\leq 2\Lambda_{\max}\left(\left(D^{-1}\right)^\top D^{-1}\right) - \frac{2}{T}\frac{\|1_T^\top u\|^2}{u^\top u} \leq 2\Lambda_{\max}\left(\left(D^{-1}\right)^\top D^{-1}\right).\end{aligned} \tag{31}$$

Since the above inequality holds for any vector u, including the eigenvector associated with the largest eigenvalue of $2A^\top A$, it follows that

$$\Lambda_{\max}\left\{2A^\top A\right\} \leq 2\Lambda_{\max}\left(\left(D^{-1}\right)^\top D^{-1}\right) \leq 2\left(\frac{\gamma^T - 1}{\gamma - 1}\right)^2. \tag{32}$$

Hence the Lipschitz constant of $\nabla f(r) = -2A^\top(\tilde{y} - Ar)$ is bounded above by $L \leq 2\left(\frac{1-\gamma^T}{1-\gamma}\right)^2$. □

E Proof of Proposition 5

We will make use of the *generalized sign*, defined as

$$\operatorname{sign}(z) = \begin{cases} -1 & \text{if } z < 0\\ [-1, 1] & \text{if } z = 0\\ 1 & \text{if } z > 0.\end{cases} \tag{33}$$

When z is a vector, the operation $\operatorname{sign}(z)$ is applied componentwise. Note that the generalized sign is the subdifferential of the ℓ_1 norm.

Proof of Proposition 5. The solution $\hat{r}$ to (15) satisfies the optimality condition

$$0 \in r - \tilde{x} + s\lambda C^T \text{sign}(Cr), \tag{34}$$

where C is a $(T-2) \times T$ matrix defined as

$$C = \begin{pmatrix} 0 & -1 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & -1 & 1 & \dots & 0 & 0 \\ \vdots & & & \ddots & \ddots & & \\ 0 & 0 & 0 & 0 & \dots & -1 & 1 \end{pmatrix}.$$

Furthermore, $\tilde{r}$ is a solution to (16) if and only if there exists some $\tilde{\mu}$ such that the pair $(\tilde{r}, \tilde{\mu})$ satisfies the Karush-Kuhn-Tucker optimality conditions (Boyd & Vandenberghe 2004), given by

$$0 \in r - \tilde{x} + s\lambda C^T \text{sign}(Cr) - \mu, \tag{35}$$
$$r_t \geq 0, \quad t = 1, \dots, t, \tag{36}$$
$$\mu_t \geq 0, \quad t = 1, \dots, t, \tag{37}$$
$$r_t \mu_t = 0, \quad t = 1, \dots, t. \tag{38}$$

To complete the proof, we will show that $\tilde{r}_t \equiv \max(\hat{r}_t, 0)$ and $\tilde{\mu}_t \equiv -\min(\hat{r}_t, 0)$ satisfy (35)–(38).

The fact that $\tilde{r}_t$ and $\tilde{\mu}_t$ satisfy (36)–(38) follows by inspection. It remains to show that $\tilde{r}_t$ and $\tilde{\mu}_t$ satisfy (35). Notice that $\hat{r}_t = \tilde{r}_t - \tilde{\mu}_t$. Therefore, because $\hat{r}$ solves (15), it follows directly from (34) that

$$0 \in \tilde{r} - \tilde{x} + s\lambda C^T \text{sign}(C\hat{r}) - \tilde{\mu}. \tag{39}$$

Inspection of the matrix C reveals that the elements of $C\hat{r}$ are of the form $\hat{r}_{t+2} - \hat{r}_{t+1}$. Furthermore, it is straightforward to show that (i) $\hat{r}_{t+2} - \hat{r}_{t+1} > 0$ implies that $\tilde{r}_{t+2} - \tilde{r}_{t+1} \geq 0$; (ii) $\hat{r}_{t+2} - \hat{r}_{t+1} < 0$ implies that $\tilde{r}_{t+2} - \tilde{r}_{t+1} \leq 0$; and (iii) $\hat{r}_{t+2} - \hat{r}_{t+1} = 0$ implies that $\tilde{r}_{t+2} - \tilde{r}_{t+1} = 0$. Therefore, $\text{sign}(C\hat{r}) \subseteq \text{sign}(C\tilde{r})$. Combining this with (39) directly implies that the pair $(\tilde{r}, \tilde{\mu})$ satisfies (35). □

F Proof of Proposition 6

Proof of Proposition 6. The Lipschitz constant of the function $-2A^\top(\tilde{y} - Ar) + 2\lambda Zr$ is given by the largest eigenvalue of $2A^\top A + 2\lambda Z$ (Bubeck et al. 2015). Since $A^\top A$ and Z are symmetric and real, the largest eigenvalue, $\Lambda_{\max}$, of their sum is bounded by the sum of their largest eigenvalues:

$$\Lambda_{\max}\left(2A^\top A + 2\lambda Z\right) \leq 2\Lambda_{\max}\left(A^\top A\right) + 2\lambda \Lambda_{\max}\left(Z\right). \tag{40}$$

.

It is not hard to show that $\Lambda_{\max}(Z) \leq 4$.

Together with (32) and (40), it follows that the Lipschitz constant L satisfies $L \leq 2\frac{\left(1-\gamma^T\right)^2 + 4\lambda(1-\gamma)^2}{(1-\gamma)^2}$. □

G Analytic Solution for the Continuously-Varying Spike-Rate Problem without Non-Negativity Constraints

We consider problem (10) with $n = 2$ and without the non-negativity rate constraint, meaning the problem

$$\text{minimize}_{r_1,\ldots,r_T} \left\{ \|\tilde{y} - Ar\|^2 + \lambda \sum_{t=3}^{T} (r_t - r_{t-1})^2 \right\}. \tag{41}$$

This problem can be solved analytically by computing $\nabla f(r) = 0$, with $f(r) = \|\tilde{y} - Ar\|^2 + \lambda r^\top Z r$ and Z defined in (21), yielding:

$$-2A^\top(\tilde{y} - Ar) + 2\lambda Z r = 0. \tag{42}$$

Solving for r requires finding a pseudo-inverse for the rank $(T-1)$ matrix $A^\top A + \lambda Z$ of size $T \times T$, which can be computationally demanding. The solution takes the form

$$r = \left(A^\top A + \lambda Z\right)^+ A^\top \tilde{y} \tag{43}$$

where $\left(A^\top A + \lambda Z\right)^+$ denotes the pseudo-inverse.

H Simulation Details

H.1 Piece-wise Constant Rate

We generated changepoints for the rate from a discrete Unif.$(0, 50)$ distribution. Between each pair of changepoints, we generated the rate from a $N(0, 10^2)$ distribution. We then made the rate trace nonnegative by subtracting out its minimum value.

We generated the underlying calcium trace $\overline{c}$ from the rate trace r by calculating $\overline{c} = D^{-1} r$, where we used $\gamma = 0.95$ in constructing D^{-1} (29). The value $\gamma = 0.95$ was estimated as the decay of the calcium in Gcamp6s across 100 milliseconds (Chen et al. 2013).

We made use of only the last $T = 600$ timepoints of the calcium and the rate traces we generated, so that $\overline{c}_1 \neq r_1$, as is the case in a real experiment in which the activity before the first measured point is unknown.

Finally, we generated the observed florescence trace $\overline{y}$ using the model (4) with variance $\bar{\sigma}^2 = 0.1\,(\max(\overline{c}) - \min(\overline{c}))$ for the noise $\overline{\epsilon}$, and a shift $\overline{\beta}_0 = -\frac{1}{T}\sum_{t=1}^{T} \overline{y}_t + \overline{\beta}_{rnd}$ where $\overline{\beta}_{rnd} \sim$ Unif. $(0, 0.1\,(\max(\overline{y}) - \min(\overline{y})))$.

H.2 Continuously-Varying Rate

We generated continuous rate traces by integrating the neural network given by the equation $\frac{dx}{dt} = -x + J\tanh(x)$, where x is a vector of size $N = 1000$, and J is an $N \times N$ matrix with i.i.d. entries $J_{ij} \sim N(0, 4/N)$. These choices led to chaotic dynamics (Sompolinsky et al. 1988). We sampled the activity of x_1 every 0.5 time steps for 600 steps (resulting in $T = 1200$ time points). To ensure non-negativity of the rate trace, we subtracted out the minimum activity.

Given the rate trace, we then repeated the steps detailed in Appendix H.1 to generate the calcium and fluorescence.

References

Aimon, S., Katsuki, T., Grosenick, L., Broxton, M., Deisseroth, K. & Greenspan, R. J. (2015), 'Activity sources from fast large-scale brain recordings in adult drosophila', *bioRxiv* p. 033803.

Allen, W. E., Kauvar, I. V., Chen, M. Z., Richman, E. B., Yang, S. J., Chan, K., Gradinaru, V., Deverman, B. E., Luo, L. & Deisseroth, K. (2017), 'Global representations of goal-directed behavior in distinct cell types of mouse neocortex', *Neuron* **94**(4), 891–907.

Boyd, S. & Vandenberghe, L. (2004), *Convex Optimization*, Cambridge University Press.

Bubeck, S. et al. (2015), 'Convex optimization: Algorithms and complexity', *Foundations and Trends® in Machine Learning* **8**(3-4), 231–357.

Chen, T.-W., Li, N., Daie, K. & Svoboda, K. (2017), 'A map of anticipatory activity in mouse motor cortex', *Neuron* **94**(4), 866–879.

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. et al. (2013), 'Ultrasensitive fluorescent proteins for imaging neuronal activity', *Nature* **499**(7458), 295.

Clancy, K. B., Orsolic, I. & Mrsic-Flogel, T. D. (2019), 'Locomotion-dependent remapping of distributed cortical networks', *Nature Neuroscience* p. 1.

Condat, L. (2013), 'A direct algorithm for 1-d total variation denoising.', *IEEE Signal Process. Lett.* **20**(11), 1054–1057.

Friedrich, J., Zhou, P. & Paninski, L. (2017), 'Fast online deconvolution of calcium imaging data', *PLoS Computational Biology* **13**(3), e1005423.

Hoefling, H. (2010), 'A path algorithm for the fused lasso signal approximator', *Journal of Computational and Graphical Statistics* **19**(4), 984–1006.

Jewell, S., Hocking, T. D., Fearnhead, P. & Witten, D. (2019), 'Fast nonconvex deconvolution of calcium imaging data', *Biostatistics* .
URL: *https://doi.org/10.1093/biostatistics/kxy083*

Jewell, S. & Witten, D. (2018), 'Exact spike train inference via ℓ_0 optimization', *The Annals of Applied Statistics* **12**(4), 2457.

Kalchenko, V., Israeli, D., Kuznetsov, Y. & Harmelin, A. (2014), 'Transcranial optical vascular imaging (tovi) of cortical hemodynamics in mouse brain', *Scientific Reports* **4**, 5839.

Lucy, L. B. (1974), 'An iterative technique for the rectification of observed distributions', *The Astronomical Journal* **79**, 745.

Makino, H., Ren, C., Liu, H., Kim, A. N., Kondapaneni, N., Liu, X., Kuzum, D. & Komiyama, T. (2017), 'Transformation of cortex-wide emergent properties during motor learning', *Neuron* **94**(4), 880–890.

Mann, K., Gallen, C. L. & Clandinin, T. R. (2017), 'Whole-brain calcium imaging reveals an intrinsic functional network in drosophila', *Current Biology* **27**(15), 2389–2396.

Masino, S., Kwon, M., Dory, Y. & Frostig, R. (1993), 'Characterization of functional organization within rat barrel cortex using intrinsic signal optical imaging through a thinned skull', *Proceedings of the National Academy of Sciences* **90**(21), 9998–10002.

Musall, S., Kaufman, M. T., Gluf, S. & Churchland, A. K. (2018), 'Movement-related activity dominates cortex during sensory-guided decision making', *BioRxiv* p. 308288.

Parikh, N., Boyd, S. et al. (2014), 'Proximal algorithms', *Foundations and Trends® in Optimization* **1**(3), 127–239.

Pnevmatikakis, E. A., Soudry, D., Gao, Y., Machado, T. A., Merel, J., Pfau, D., Reardon, T., Mu, Y., Lacefield, C., Yang, W. et al. (2016), 'Simultaneous denoising, deconvolution, and demixing of calcium imaging data', *Neuron* **89**(2), 285–299.

Richardson, W. H. (1972), 'Bayesian-based iterative method of image restoration', *JOSA* **62**(1), 55–59.

Silasi, G., Xiao, D., Vanni, M. P., Chen, A. C. & Murphy, T. H. (2016), 'Intact skull chronic windows for mesoscopic wide-field imaging in awake mice', *Journal of Neuroscience Methods* **267**, 141–149.

Sompolinsky, H., Crisanti, A. & Sommers, H.-J. (1988), 'Chaos in random neural networks', *Physical Review Letters* **61**(3), 259.

Swanson, R., Basu, J. & Buzsaki, G. (2018), 'Investigating hippocampo-cortical dialogue using wide-field calcium imaging and electrophysiology in vivo', *Soc. For Neuroscience Meeting* .

Tibshirani, R. (1996), 'Regression shrinkage and selection via the lasso', *Journal of the Royal Statistical Society: Series B (Methodological)* **58**(1), 267–288.

Tibshirani, R. J., Taylor, J. et al. (2012), 'Degrees of freedom in lasso problems', *The Annals of Statistics* **40**(2), 1198–1232.

Tibshirani, R., Saunders, M., Rosset, S., Zhu, J. & Knight, K. (2005), 'Sparsity and smoothness via the fused lasso', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **67**(1), 91–108.

Vogelstein, J. T., Watson, B. O., Packer, A. M., Yuste, R., Jedynak, B. & Paninski, L. (2009), 'Spike inference from calcium imaging using sequential monte carlo methods', *Biophysical Journal* **97**(2), 636–655.

Wekselblatt, J. B., Flister, E. D., Piscopo, D. M. & Niell, C. M. (2016), 'Large-scale imaging of cortical dynamics during sensory perception and behavior', *American Journal of Physiology-Heart and Circulatory Physiology* .