AN EXTENSION OF THE FRI FRAMEWORK FOR CALCIUM TRANSIENT DETECTION

Stephanie Reynolds [1,3] *Caroline S. Copeland* [2,3] *Simon R. Schultz* [2,3] *Pier Luigi Dragotti* [1,3]

[1] Department of Electrical and Electronic Engineering, [2] Department of Bioengineering,
[3] Centre for Neurotechnology, Imperial College London

ABSTRACT

Two-photon calcium imaging of the brain allows the spatiotemporal activity of neuronal networks to be monitored at cellular resolution. In order to analyse this activity it must first be possible to detect, with high temporal resolution, spikes from the time series corresponding to single neurons. Previous work has shown that finite rate of innovation (FRI) theory can be used to reconstruct spike trains from noisy calcium imaging data. In this paper we extend the FRI framework for spike detection from calcium imaging data to encompass data generated by a larger class of calcium indicators, including the genetically encoded indicator GCaMP6s. Furthermore, we implement least squares model-order estimation and perform a noise reduction procedure ('pre-whitening') in order to increase the robustness of the algorithm. We demonstrate high spike detection performance on real data generated by GCaMP6s, detecting 90% of electrophysiologically-validated spikes.

Index Terms— Calcium imaging, Calcium transient detection, Finite rate of innovation, GCaMP6s

1. INTRODUCTION

Optical imaging of populations of neurons at cellular resolution may prove crucial to developing our understanding of the function of the brain. In order to analyse the spatiotemporal activity of neuronal networks, one must first be able to detect, with high temporal precision, the time at which action potentials (spikes) were fired from individual neurons.

As the concentration of intracellular free calcium is a reliable indicator of spiking activity, several optical imaging methods rely upon calcium-sensitive fluorescent indicators (calcium indicators) which visualise spiking activity via a change in their fluorescence intensity. Recent advances in protein engineering have produced genetically encoded calcium indicators (GECIs) which have, for the first time, exceeded the sensitivity of the traditionally used synthetic indicators [1]. GECIs have a proven capability for imaging the same *in vivo* neuronal populations over multiple weeks and they can be targeted to selected cellular and subcellular compartments. Due to the above advantages of GECIs, they are becoming the preferred tool for calcium imaging experiments. Spike detection algorithms which are suited to the kinetics of these indicators are thus required.

A spike in a neuron produces a pulse with a characteristic shape in that neuron's fluorescence signal, this pulse is referred to as a calcium transient. Several algorithms employ template-matching approaches which locate portions of the fluorescence signal that correspond to the expected pulse template [2, 3]. The performance of such algorithms which include strict assumptions on the template's amplitude [3] deteriorates on real data, in which transient amplitudes vary greatly. Vogelstein et al. developed a fast algorithm that performs a *maximum a posteriori* estimation to infer the most likely spike train given the imaging data and a model of intracellular calcium dynamics [4].

In [5], Oñativia et al. exploited the fact that calcium imaging data, which can be modelled as streams of decaying exponentials, are a class of signals with a finite rate of innovation (FRI) [6]. They used FRI theory to develop a fast spike detection algorithm which demonstrated both high accuracy and high temporal precision when detecting spikes from calcium imaging data generated by the synthetic dye Oregon Green BAPTA 1-AM (OGB-1).

The main focus of this paper is to extend the FRI framework for calcium transient detection to be used on a larger class of calcium indicators. Oñativia et al. modelled the characteristic pulse shape as an instantaneous rise and an exponential decay. This is a good approximation for some calcium indicators, but not for those with a slow rise such as the GECI GCaMP6s, which has already been shown to be highly useful for the study of neuronal networks [1]. Transients generated by GCaMP6s take 200-300ms to reach peak amplitude and thus will be detected with low temporal precision by algorithms which assume an instantaneous rise time.

In Section 2 we generalise the FRI framework for calcium transient detection for a pulse template which approximates the dynamics of a larger class of calcium indicators. In Section 2.2 we introduce a method to increase the robustness of spike detection from noisy fluorescence signals and in Section 2.3 we outline our least squares model-order estimation framework. Finally, in Section 3 we demonstrate the performance of the modified FRI algorithm on real data generated by the calcium indicator GCaMP6s.

978-1-4799-2349-6/16/$31.00

2. FINITE RATE OF INNOVATION THEORY APPLIED TO CALCIUM TRANSIENT DETECTION

A spike in a neuron produces a calcium transient with a characteristic pulse shape in the corresponding neuron's fluorescence signal. This signal can therefore be modelled as a convolution of the spike train $x(t) = \sum_{k=1}^{K} a_k \delta(t - t_k)$ with the known characteristic pulse shape $p(t)$, such that

$$f(t) = x(t) * p(t). \tag{1}$$

Finite Rate of Innovation (FRI) theory is a framework for the sampling and reconstruction of signals that can be completely defined by a finite number of free parameters. The signal $f(t)$ is an example of such a signal as it is completely defined by the parameter set $\{a_k, t_k\}_{k=1}^{K}$.

In [5], Oñativia et al. developed an FRI algorithm to detect the time points of calcium transients whose shape is characterised by an instantaneous rise and an exponential decay. They initially filter the signal $f(t)$ with an exponential reproducing kernel and compute weighted finite differences of the samples. These operations allow the authors to transform the problem from one of estimating the time points of calcium transients to the classical FRI problem of retrieving the locations of a stream of Diracs. Transforming the problem into a classical one in the FRI framework enables the authors to use FRI methods (see [7]) to retrieve the time points of the calcium transients.

We now focus on extending the FRI framework for calcium transient detection to those whose shape is characterised by a slower (not instantaneous) rise and an exponential decay. We model their characteristic pulse shape as

$$p(t) = \left(e^{-\alpha t} - e^{-\gamma t}\right) \mathbb{1}_{t \geq 0}, \tag{2}$$

where α and γ are known parameters. In order to use the FRI framework for this pulse shape, we must first identify the filtering scheme that transforms the estimation problem into sampling and reconstructing a stream of Diracs.

Proposition 1. *Filtering $f(t) = x(t) * p(t)$ with the scheme in* (3)

$$\begin{aligned} y_n &= \left\langle f(t), \varphi\left(\tfrac{t}{T} - n\right) \right\rangle \\ z_n &= y_n - e^{-\alpha T} y_{n-1} \\ w_n &= z_n - e^{-\gamma T} z_{n-1}, \end{aligned} \tag{3}$$

is analogous to filtering $x(t)$ with the kernel $\psi(-\frac{t}{T})$, where

$$\psi(t) = \varphi(t) * \beta_{-\alpha T}(-t) * \beta_{-\gamma T}(-t), \tag{4}$$

$\beta_{-\alpha T}$, $\beta_{-\gamma T}$ are first order E-splines and T is the sampling period.

Proof. From the initial filtering operation we obtain

$$\begin{aligned} y_n &= \left\langle f(t), \varphi\left(\tfrac{t}{T} - n\right) \right\rangle \\ &= \left\langle x(t) * p(t), \varphi\left(\tfrac{t}{T} - n\right) \right\rangle \\ &= \left\langle x(t), p(-t) * \varphi\left(\tfrac{t}{T} - n\right) \right\rangle. \end{aligned} \tag{5}$$

We take finite differences as in Eq. (3) so that we have

$$w_n = y_n - \left(e^{-\alpha T} + e^{-\gamma T}\right) y_{n-1} + e^{-\alpha T} e^{-\gamma T} y_{n-2}. \tag{6}$$

We write $w_n = \langle x(t), h(t) \rangle$ where, from (5) and the linearity of the inner product, we have

$$\begin{aligned} h(t) =& p(-t) * \varphi\left(\tfrac{t}{T} - n\right) \\ & - p(-t) * \varphi\left(\tfrac{t}{T} - (n-1)\right)\left(e^{-\alpha T} + e^{-\gamma T}\right) \\ & + p(-t) * \varphi\left(\tfrac{t}{T} - (n-2)\right) e^{-\alpha T} e^{-\gamma T}. \end{aligned} \tag{7}$$

Due to Parseval's relation w_n can be expressed as

$$w_n = \tfrac{1}{2\pi} \left\langle \mathscr{F}\left\{x(t)\right\}, \mathscr{F}\left\{h(t)\right\} \right\rangle, \tag{8}$$

where $\mathscr{F}\left\{x(t)\right\} := \hat{x}(\omega)$ denotes the Fourier Transform (FT) of $x(t)$. Noting that

$$\mathscr{F}\left\{p(-t)\right\} = \tfrac{\gamma - \alpha}{(\alpha - i\omega)(\gamma - i\omega)}, \tag{9}$$

$$\mathscr{F}\left\{\varphi\left(\tfrac{t}{T} - n\right)\right\} = T\hat{\varphi}(\omega T) e^{-iwnT}, \tag{10}$$

we have

$$\begin{aligned} &\mathscr{F}\left\{h(t)\right\} = \\ &T\hat{\varphi}(\omega T) e^{-iwnT} \tfrac{\left(1 - e^{-T(\alpha - i\omega)}\right)\left(1 - e^{-T(\gamma - i\omega)}\right)}{(\alpha - i\omega)(\gamma - i\omega)} (\gamma - \alpha). \end{aligned} \tag{11}$$

The FT of a time-reversed and scaled E-spline with parameter $-\alpha T$ is

$$\mathscr{F}\left\{\beta_{-\alpha T}\left(-\tfrac{t}{T}\right)\right\} = \tfrac{1 - e^{-T(\alpha - i\omega)}}{\alpha - i\omega}. \tag{12}$$

From (10) and (12) it follows that

$$\begin{aligned} \mathscr{F}\left\{h(t)\right\} =& (\gamma - \alpha)\, \mathscr{F}\left\{\varphi\left(\tfrac{t}{T} - n\right)\right\} \\ & \mathscr{F}\left\{\beta_{-\alpha T}\left(-\tfrac{t}{T}\right)\right\} \mathscr{F}\left\{\beta_{-\gamma T}\left(-\tfrac{t}{T}\right)\right\}. \end{aligned} \tag{13}$$

Using Parseval's relation once more we can write

$$\begin{aligned} w_n =& (\gamma - \alpha) \langle x(t), \\ & \varphi\left(\tfrac{t}{T} - n\right) * \beta_{-\alpha T}\left(-\tfrac{t}{T}\right) * \beta_{-\gamma T}\left(-\tfrac{t}{T}\right) \rangle, \end{aligned} \tag{14}$$

which, with $\psi(t) = \varphi(t) * \beta_{-\alpha T}(-t) * \beta_{-\gamma T}(-t)$, is the statement of the proposition. □

2.1. Detection of calcium transients in the noisy scenario

The process of detecting the time points of calcium transients from a noisy fluorescence signal is described in further detail in this section and stated in Algorithm 1. We model the samples $\tilde{y}_n$ as being corrupted by additive white noise so that we have

$$\tilde{y}_n = \left\langle f(t), \varphi\left(\tfrac{t}{T} - n\right) \right\rangle + \epsilon_n = y_n + \epsilon_n, \tag{15}$$

where ϵ_n are i.i.d Gaussian with zero mean and standard deviation σ. The sampling kernel φ is chosen to be an exponential reproducing kernel. This is defined such that, when it is combined in a weighted sum with shifted versions of itself, it reproduces exponentials:

$$\sum_{n \in \mathbb{Z}} c_{m,n} \varphi(t - n) = e^{\theta_m t}. \tag{16}$$

Algorithm 1: Estimate spike times and amplitudes

Input: $f(t), \quad K, \quad \alpha, \quad \gamma$
Output: $\{\hat{a}_j, \hat{t}_j\}_{j=1}^{K}$

1 Filter: $\tilde{y}_n = \left\langle f(t), \varphi\left(\frac{t}{T} - n\right)\right\rangle + \epsilon_n$
2 Weighted finite differences: $\tilde{z}_n = \tilde{y}_n - e^{-\alpha T}\tilde{y}_{n-1}$
3 Weighted finite differences: $\tilde{w}_n = \tilde{z}_n - e^{-\gamma T}\tilde{z}_{n-1}$
4 Compute sample moments: $\tilde{s}_m = \sum_n d_{m,n}\tilde{w}_n$
5 Create Toeplitz matrix $\tilde{\mathbf{S}}$ from $\tilde{s}_m$
6 Pre-whiten Toeplitz matrix: $\tilde{\mathbf{S}}' = \tilde{\mathbf{S}}\mathbf{W}$
7 Use Matrix Pencil Method to estimate $\{\hat{t}_j\}_{j=1}^{K}$ from $\tilde{\mathbf{S}}'$
8 Estimate $\{\hat{a}_j\}_{j=1}^{K}$ by least squares from samples and resynthesized sample estimates

We refer to $c_{m,n}$ as the coefficients of the kernel. The weighted finite differences in (3) result in noisy samples $\tilde{w}_n = w_n + \vartheta_n$, where ϑ_n are Gaussian but no longer i.i.d. We compute sample moments

$$\begin{aligned}\tilde{s}_m &= \sum_n d_{m,n}\tilde{w}_n = \sum_n d_{m,n}w_n + \sum_n d_{m,n}\vartheta_n \\ &= s_m + b_m,\end{aligned} \tag{17}$$

using the coefficients $d_{m,n}$ of the kernel ψ (Eq. (4)) which also reproduces exponentials. By choosing φ so that ψ reproduces exponents in the form $\zeta_m = \zeta_0 + m\lambda$, we can write s_m in power-sum series form:

$$s_m = \sum_{k=1}^{K} l_k u_k^m \tag{18}$$

where $l_k = (\gamma - \alpha)a_k e^{\zeta_0 \frac{t_k}{T}}$ and $u_k = e^{\lambda \frac{t_k}{T}}$. The sample moments s_m are used to construct a Toeplitz matrix $\tilde{\mathbf{S}} = \mathbf{S} + \mathbf{B}$, where $\mathbf{S}$ is the idealised noiseless Toeplitz matrix and $\mathbf{B}$ is due to the noise b_m corrupting the sample moments. The parameters u_k and thus the time points of the calcium transients t_k are then retrieved from $\tilde{\mathbf{S}}$ by the matrix pencil method [8]. As the noise has been colored by the filtering process, we first include a pre-whitening step which improves the robustness of the subspace estimation from $\tilde{\mathbf{S}}$.

2.2. Pre-whitening to increase robustness to noise

We implement the matrix pencil method to recover the parameters u_k, and thus $t_k = \frac{T}{\lambda}\ln(u_k)$, from the noisy Toeplitz matrix $\tilde{\mathbf{S}} = \mathbf{S} + \mathbf{B}$. This method performs well when the matrix $\mathbf{B}$ is white, i.e. $\mathbf{R}_B := \mathrm{E}\left[\mathbf{B}^H\mathbf{B}\right] = a\mathbf{I}$ for $a \in \mathbb{R}$.

The finite differences in Eq. (3) result in colored noise, such that $\mathbf{R}_B \neq a\mathbf{I}$. We therefore follow the method of Urigüen et al. [10] to pre-whiten $\tilde{\mathbf{S}}$. We do this by post-multiplying $\tilde{\mathbf{S}}$ with the matrix $\mathbf{W} := \mathbf{R}_B^{\dagger/2}$ (where $^{\dagger/2}$ denotes the square root of the pseudo-inverse) to obtain

$$\tilde{\mathbf{S}}' := \mathbf{S}' + \mathbf{B}' = \mathbf{S}\mathbf{W} + \mathbf{B}\mathbf{W} = \tilde{\mathbf{S}}\mathbf{W}. \tag{19}$$

We thus have a noisy Toeplitz matrix $\tilde{\mathbf{S}}'$ which is corrupted by white noise. The matrix pencil method can then be applied to $\tilde{\mathbf{S}}'$ to obtain the same signal parameters u_k as would be obtained from $\tilde{\mathbf{S}}$ whilst maintaining robustness to noise.

2.3. Least squares model-order estimation

We implement a least squares model-order estimation framework similar to that proposed by Doğan et al. [11] to estimate the number of spikes in a window of the trace. For each possible model order k we estimate the corresponding spike times according to Algorithm 1 and compute the training error between the samples and the resynthesized sample estimates. We then estimate the model order $\hat{k}$ as the value of k which minimises the training error. The corresponding spike times are stored. This procedure is completed sequentially in a sliding window along the fluorescence signal. Finally, spike time estimates which are not consistently detected across windows are deemed to be spurious spikes due to noise and are pruned.

3. RESULTS

We now demonstrate the performance of the modified FRI algorithm on real GCaMP6s imaging data [9] (for experimental methods see [1]). The dataset, which is recorded from *in vivo* mouse V1 neurons, contains simultaneous calcium imaging (sampled at 60Hz) and electrophysiology. This allows us to compare estimated spike positions from the calcium imaging data with the electrophysiological ground truth. We assess algorithm performance on 3 traces of combined length 678s containing 532 spikes.

A spike is deemed detected if an estimate is within 0.033s (2 sample widths) of the real spike, in this case we denote the estimate as a true positive. As the rise of a GCaMP6s transient lasts approximately 15 sample widths, this is a strict target. An estimate is deemed to be a false positive if it is not within 2 samples of a real spike.

The modified FRI algorithm correctly detects over 90% of electrophysiologically validated spikes (Fig. 1). Furthermore, as is shown in Fig. 1c, the modified FRI algorithm correctly detects a higher proportion of spikes on all three datasets than both the original FRI algorithm and Vogelstein et al.'s fast non-negative deconvolution algorithm [4].

Vogelstein et al.'s method detects a high proportion of false positives on this dataset. This is likely due to their assumption of a uniform fluorescence change per spike which, as is noted in [1], is not true for GCaMP6s. Although the deconvolution algorithm has previously shown good performance on *in vitro* data generated by the synthetic dye OGB-1, it doesn't perform as well on GCaMP6s data. This demonstrates the necessity of tailoring algorithms to the calcium indicator when performing calcium transient detection.

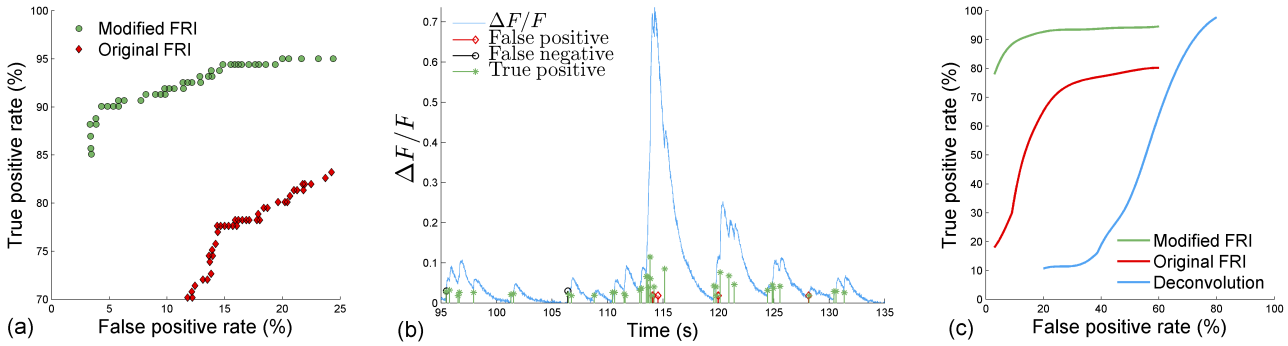

Fig. 1: The modified FRI algorithm outperforms the original FRI algorithm and Vogelstein et al.'s algorithm [4] on three datasets for which electrophysiological ground truth is available [9]. A spike is deemed detected if one is estimated within 0.033s (2 samples) of the true spike position. a) ROC curves on a dataset of length 239s containing 181 spikes. b) The modified FRI algorithm's spike detection performance on a section of the trace. c) ROC curves averaged over 3 datasets (total length 678s, 532 spikes). To compute the average curves per algorithm we averaged the least squares spline fit to each ROC curve.

4. CONCLUSION

We extended the FRI framework for spike detection from calcium imaging data to encompass calcium transients with a slow rise, such as those generated by the genetically encoded calcium indicator GCaMP6s. We introduced a noise reduction technique (pre-whitening) and least squares model-order estimation to improve the robustness of the algorithm. On real GCaMP6s data we showed that these modifications increased the spike detection rate of the algorithm for all false positive rates. Furthermore, on real data we achieve spike detection rates of 90% of electrophysiologically-validated spikes within 2 sample widths (0.033s) of the real spike.

5. REFERENCES

[1] T.-W. Chen et al., "Ultrasensitive fluorescent proteins for imaging neuronal activity," *Nature*, vol. 499, no. 7458, pp. 295–300, 2013.

[2] S. R. Schultz, K. Kitamura, A. Post-Uiterweer, J. Krupic, and M. Häusser, "Spatial pattern coding of sensory information by climbing fiber-evoked calcium signals in networks of neighboring cerebellar purkinje cells," *Journal of Neuroscience*, vol. 29, no. 25, pp. 8005–8015, 2009.

[3] B. F. Grewe, D. Langer, H. Kasper, B. M. Kampa, and F. Helmchen, "High-speed in vivo calcium imaging reveals neuronal network activity with near-millisecond precision.," *Nature methods*, vol. 7, no. 5, pp. 399–405, 2010.

[4] J. T. Vogelstein, A. M. Packer, T. A. Machado, T. Sippy, B. Babadi, R. Yuste, and L. Paninski, "Fast nonnegative deconvolution for spike train inference from population calcium imaging," *Journal of Neurophysiology*, vol. 104, no. 6, pp. 3691–3704, 2010.

[5] J. Oñativia, S. R. Schultz, and P. L. Dragotti, "A finite rate of innovation algorithm for fast and accurate spike detection from two-photon calcium imaging," *Journal of Neural Engineering*, vol. 10, no. 4, pp. 046017, 2013.

[6] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *Signal Processing, IEEE Transactions on*, vol. 50, no. 6, pp. 1417–1428, 2002.

[7] T. Blu, P. L. Dragotti, M. Vetterli, P. Marziliano, and L. Coulot, "Sparse sampling of signal innovations," *Signal Processing Magazine, IEEE*, vol. 25, no. 2, pp. 31–40, 2008.

[8] Y. Hua and T. K. Sarkar, "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 38, no. 5, pp. 814–824, 1990.

[9] "Simultaneous imaging and loose-seal cell-attached electrical recordings from neurons expressing a variety of genetically encoded calcium indicators," GENIE project, Janelia Farm Campus, HHMI; Karel Svoboda (contact). (2015). CRCNS.org. http://dx.doi.org/10.6080/K02R3PMN.

[10] J. A. Urigüen, T. Blu, and P. L. Dragotti, "FRI sampling with arbitrary kernels," *Signal Processing, IEEE Transactions on*, vol. 61, no. 21, pp. 5310–5323, 2013.

[11] Z. Doğan, C. Gilliam, T. Blu, and D. Van De Ville, "Reconstruction of finite rate of innovation signals with model-fitting approach," *Signal Processing, IEEE Transactions on*, vol. 63, no. 22, pp. 6024–6036, 2015.