

A Robust and Efficient Method to Recover Neural Events from Noisy and Corrupted Data

Eva L. Dyer, Christoph Studer, Jacob T. Robinson, and Richard G. Baraniuk

Dept. of Electrical & Computer Engineering, Rice University
e-mail: {e.dyer, studer, jtrobins, richb}@rice.edu

Abstract—In a variety of neural data analysis problems, “neural events” such as action potentials or post-synaptic potentials (PSPs), must be recovered from noisy and possibly corrupted measurements. For instance, in calcium imaging, an action potential or group of action potentials give rise to a stereotyped calcium signal with quick rise and slow decay. In this work, we develop a general-purpose method for *learning a template waveform* that characterizes the waveform of neural events and *neural event recovery* to determine the times at which such events occur. Our approach is based upon solving a *sparse signal separation* problem to separate the neural signal of interest from any noise and other corruptions that arise due to baseline drift, measurement noise, and breathing/motion artifacts. For both synthetic and real data, we demonstrate that our approach accurately detects neural events and learns the underlying template waveform, even in the presence of strong amounts of noise and corruptions. The method’s robustness and simplicity makes it amenable for use in the analysis of datasets arising in large-scale studies of both time-varying calcium imaging and whole cell electrophysiology.

I. INTRODUCTION

Experimental neuroscience has experienced dramatic growth over the past few years. Out of this growth, experimentalists now possess the tools necessary to modulate and record from increasingly large populations of neurons [1]. In order to analyze the data generated from large-scale experiments, computationally efficient, robust, and automated methods for neural data analysis are of paramount importance.

While a great deal of work has focused on the detection and classification of spikes (the basic unit of information in neural systems) from both single and multi-electrode arrays, researchers are increasingly turning toward optical imaging techniques that have the potential to record from a greater number of individual neurons. The fact that these methods (voltage or calcium-sensitive imaging) rely on indirect optical measurements of the transmembrane voltage, the resulting estimates of spike timing are often imprecise. Further complicating things is the fact that the signal generated from a sequence of spikes is often corrupted by noise and/or an additive baseline or drift component that can vary in its structure, depending on the experimental setup and conditions.

In many settings, neural signals that are associated with either the action potentials (APs) of the cell-under-test or the neural events generated by the pre-synaptic partners of the cell-under-test (post synaptic potentials), give rise to a stereotyped waveform that can be modeled by a common *template* waveform that approximates the time-course of the neural signal produced by a single spike event. Thus, a

common method to extract the times at which events occur, is to use a matched filter to correlate the known template with the measured signal [2]. If the correlation between the template and the signal exceeds a certain threshold, then such algorithms declare the presence of a spike event. Previous studies of event detection in both the analysis of PSPs and calcium imaging data, have demonstrated that standard template matching-based approaches that rely on the correlation between the observations are unable to resolve events that occur in rapid succession or appear in noisy conditions [3], [4]. Hence, *deconvolution* methods that aim to reverse the effects of convolution on the data, are employed to improve the temporal resolution of correlation-based methods.

In this work, we develop a novel, robust method for *neural event detection* that improves upon state-of-the-art deconvolution-based methods such as [5], [3] in a number of ways. First, we introduce a novel sparsity-based deconvolution approach that not only finds an estimate of the time and amplitude of neural events but also separates it from the baseline and noise components corrupting the measurement of the neural activity. By exploiting known signal structure in both the neuronal signal and the baseline component, we demonstrate that the proposed approach reliably recovers the neuronal signal from significant amounts of baseline drift and noise, in both real and synthetic data. The second contribution is the development of an alternating minimization framework for obtaining estimates of the underlying template waveform that generated the observed signal. This template learning method employs a constrained least-squares (LS) method to learn a kernel without any assumed parametric model; instead, we simply enforce the fact that the template must be non-negative and of limited duration relative to the length of the measurement. Our results on both synthetic and real data demonstrate that the proposed approach provides a nearly parameter-free method that enables accurate and robust event recovery from large-scale experimental datasets.

II. SIGNAL MODEL AND SPARSE SIGNAL RECOVERY

We start by introducing a convolution-based signal model for neuronal signals arising in calcium imaging and electrophysiology recordings. We then show how this model can be used to derive a robust and computationally efficient deconvolution approach to detect neural events.

A. Signal Model

In a wide range of neural datasets, the precise timing of a neural event is obscured by the fact that the event is convolved with a template waveform that produces the effective neural signal in the measurement. For this reason, sophisticated methods for event detection must be employed in place of conventional thresholding strategies used in spike detection from electrode arrays [5]. Here, we consider a general signal model for neural datasets that admit convolutional structure. We next provide a brief description of two exemplar datasets and outline the challenges that arise in the recovery and detection of neural events from these datasets.

1) *Spike recovery from calcium imaging data*: To monitor the spiking activity of large neuronal populations, calcium-sensitive fluorophores can be used to measure the intracellular calcium concentration. When an AP is emitted, this results in a transient increase in the intracellular calcium concentration that can be monitored as a change in the calcium-sensitive fluorescence. The corresponding fluorescence waveform is often modeled as an instantaneous rise followed by an exponential or multi-exponential decay curve. The template that approximates this event-driven activity is governed by the binding kinetics of the calcium-sensitive dye used to monitor calcium activity. For this reason, different neurons will produce a similar fluorescence waveform for each AP. In addition to this event-driven signal, in vivo imaging data is also corrupted by baseline drift and motion/breathing artifacts.

2) *Recovery of PSPs*: In order to study properties of synapses, plasticity, and functional connectivity of neural circuits, patch clamp techniques are employed to record post-synaptic potentials (PSPs) of a neuron-under-test, in response to stimulation of single or multiple pre-synaptic partners. Low-pass filtering by the dendritic arbor, and the time constant of the cell membrane lead to PSP waveforms that are roughly approximated by a function with a fast rise time and much slower decay. Since the amplitude of a PSP provides information about the strength of the synaptic connection between a pre-synaptic cell and the neuron-under-test, in addition to recovering the times that events occur, we must also recover the amplitude of the event.

In both of the aforementioned examples, the transformation from a neural event to the measured signal can be modeled as a *convolution* of the original signal with a template. Thus, the generation of (noiseless and uncorrupted) neuronal signals $\mathbf{y}_a \in \mathbb{R}^N$ can be described by the following signal model, $\mathbf{y}_a = \sum_{i=1}^k [\mathbf{x}_a]_i (\delta_{n_i} \otimes \mathbf{h})$, where $[\mathbf{x}_a]_i \in \mathbb{R}$ is the amplitude of the i^{th} event, $\delta_{n_i} \in \mathbb{R}^N$ is the Kronecker delta function with 1 at the n_i^{th} entry and zeros otherwise, $\mathbf{h} \in \mathbb{R}^N$ is the template waveform that represents the transformation of a spike into the domain of the observations, and k is the number of events present in \mathbf{y}_a . To simplify notation, we will rewrite \mathbf{y}_a as a linear transformation, $\mathbf{y}_a = \mathbf{A}\mathbf{x}_a$, where the amplitude of the i^{th} event, $[\mathbf{x}_a]_i$ is contained in the i^{th} entry of the vector $\mathbf{x}_a \in \mathbb{R}^N$ and the i^{th} column of \mathbf{A} consists of the template \mathbf{h} circularly shifted by i samples; thus, the matrix \mathbf{A} is circulant. Note that we have $\mathbf{A} = \mathbf{F}^* \mathbf{\Lambda} \mathbf{F}$ with \mathbf{F} being the discrete Fourier transform (DFT) matrix and $\mathbf{\Lambda} = \text{diag}(\boldsymbol{\lambda})$,

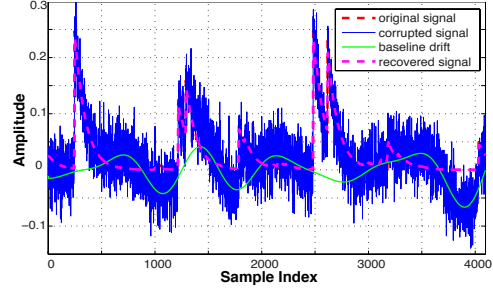


Fig. 1. Synthetic model of calcium imaging data with slowly-varying baseline component and additive noise: observations (blue), baseline component (green, solid), neural signal (red, dashed), and recovered signal (magenta, dashed).

where $\boldsymbol{\lambda} \in \mathbb{C}^N$ is the frequency-domain representation of the template waveform \mathbf{h} , i.e., $\mathbf{F}\mathbf{h} = \boldsymbol{\lambda}$.

The final model for measured neural data in $\mathbf{z} \in \mathbb{R}^N$ is given by the following input-output relation:

$$\mathbf{z} = \mathbf{y}_a + \mathbf{y}_b + \mathbf{n}. \quad (1)$$

Here, the measurement vector \mathbf{z} is composed from three components: the observed neuronal signal $\mathbf{y}_a = \mathbf{A}\mathbf{x}_a$, a correlative baseline or drift component $\mathbf{y}_b \in \mathbb{R}^N$ generated from other neural and/or measurement processes that we wish to remove from the neuronal signal \mathbf{y}_a , and additive measurement noise $\mathbf{n} \in \mathbb{R}^N$. We provide a synthetic example of calcium imaging data generated according to this model in Fig. 1; in Fig. 2 we show an example of a real intracellular measurement that is well-approximated by the model in (1).

B. Sparse Deconvolution Methods for Event Detection

State-of-the-art methods for event recovery from measurements generated in accordance with (1) use an estimate of the template \mathbf{h} to *deconvolve* the measurements to obtain an estimate of the signal \mathbf{x}_a . LS-based methods for deconvolution provide reasonable estimates of \mathbf{x}_a when the signal-to-noise ratio is high and the amount of corruptions is very low. However, when the measurements are corrupted, additional regularization or signal-specific information regarding both \mathbf{y}_a and \mathbf{y}_b is required to obtain an accurate estimate of the neural signal of interest.

In order to mitigate the drawbacks of LS-based recovery, a *sparse signal model* can be imposed on the vector \mathbf{x}_a , which assumes that the number of events k is small compared to the length of the observation, i.e., $k \ll N$. Sparsity has recently been shown to provide a powerful signal model for neural data, and sparse deconvolution techniques have been applied successfully to estimate timing information from calcium signals [5], [4] and from multi-electrode arrays [6]. To obtain a sparse estimate \mathbf{x}_a , one typically uses an ℓ_1 -norm penalty term, which leads to the well-known basis pursuit denoising (BPDN) optimization problem [7]

$$\text{(BPDN)} \quad \underset{\mathbf{x} \in \mathbb{R}^N}{\text{minimize}} \quad \|\mathbf{x}\|_1 \quad \text{subject to} \quad \|\mathbf{z} - \mathbf{A}\mathbf{x}\|_2 \leq \delta.$$

Here, the parameter $\delta \geq 0$ must be set in accordance to the amount of noise present in the system. In [4], it was shown that such a sparsity-based approach provides significant

improvements over standard deconvolution methods for neural event recovery [3].

C. Sparse Signal Separation

While sparse deconvolution techniques provide significant improvements over LS-based methods, the presence of corruptions by, e.g., a baseline drift \mathbf{y}_b , inhibits accurate estimates of \mathbf{x}_a when solving the BPDN problem. Instead, to simultaneously extract the baseline \mathbf{y}_b and the neural signal \mathbf{y}_a from the noisy measurements \mathbf{z} , a *sparse signal separation* problem can be solved. This problem has been studied extensively in the context of signal restoration and recovery from sparsely corrupted measurements, and solves [8], [9]

$$\text{(BP-SEP)} \quad \begin{cases} \text{minimize} & \|\mathbf{x}_a\|_1 + \|\mathbf{x}_b\|_1 \\ \mathbf{x}_a, \mathbf{x}_b \in \mathbb{R}^N & \\ \text{subject to} & \|\mathbf{z} - \mathbf{A}\mathbf{x}_a + \mathbf{B}\mathbf{x}_b\|_2 \leq \delta. \end{cases} \quad (2)$$

Here, we assume that the baseline \mathbf{y}_b admits a sparse representation with the respect to the basis $\mathbf{B} \in \mathbb{R}^{N \times M}$. Provided that both \mathbf{y}_a and \mathbf{y}_b admit sparse representations with respect to \mathbf{A} and \mathbf{B} , respectively, and that \mathbf{A} and \mathbf{B} are incoherent (the maximum inner product between the columns of \mathbf{A} and \mathbf{B} is small), then stable signal separation is guaranteed (see [8], [9] for the details). In what follows, we show how sparse signal separation approach can be utilized for neural event detection from noisy and corrupted measurements.

III. NERDS: NEURAL EVENT RECONSTRUCTION AND DETECTION VIA SPARSITY

We now detail our method for joint neural event detection and template estimation.

A. Algorithm Outline

In order to both recover the time instances and amplitudes of neural events from the measurements in \mathbf{z} and to estimate the underlying kernel \mathbf{h} , we propose an alternating optimization approach. Specifically, we perform the following three-step procedure: 1) we hold \mathbf{h} fixed while optimizing \mathbf{x}_a , 2) we solve a model-based thresholding operation to clean up the sparse representation \mathbf{x}_a by keeping only the dominant neural events, and 3) we hold the cleaned version $\hat{\mathbf{x}}_a$ fixed and estimate the template waveform \mathbf{h} . This procedure is repeated until both estimates converge; the associated details are provided below.

B. Step 1: Coefficient Estimation via Sparse Signal Separation

To obtain a sparse estimate of \mathbf{x}_a , while assuming that the template \mathbf{h} is fixed, we propose to solve the sparse signal separation problem provided in (2). In our application, we set \mathbf{B} to an orthonormal discrete cosine transform (DCT) matrix, which allows us to separate the baseline signal $\mathbf{y}_b = \mathbf{B}\mathbf{x}_b$ from the signal of interest \mathbf{y}_a present in the measurements \mathbf{z} . The DCT basis provides sparse representations for slowly-varying and oscillatory baseline components observed in real datasets, while also satisfying the incoherence properties with the circulant matrix \mathbf{A} (which consist of circular shifts of the template \mathbf{h}) required to enable signal separation (see [8],

[9] for the details).¹ Hence, solving BP-SEP simultaneously estimates the sparse representations \mathbf{x}_a and \mathbf{x}_b from \mathbf{z} .

C. Step 2: Peak Detection via Model-Based Thresholding

In order to suppress spurious non-zero coefficients in \mathbf{x}_a that are identified close to the true firing event of a neuron, we employ a greedy method developed for model-based sparse recovery of neural signals in [10], [4]. The deployed method accurately detects the main firing events, while suppressing non-zeros coefficients that are in the vicinity of $\Delta \in \mathbb{N}^+$ around the main peak. The window size Δ can be chosen in accordance with the refractory period of the neuron. We emphasize that this peak detection step is particularly useful in the presence of a model mismatch between the true and estimated template waveform \mathbf{h} , which often causes non-zero coefficients close to the main peak in \mathbf{x}_a . Let $\hat{\mathbf{x}}_a$ denote the coefficient vector after this cleanup step.

D. Step 3: Template Estimation via Constrained LS

The third step computes an improved estimate of the template waveform \mathbf{h} , where $\hat{\mathbf{x}}_a$ is assumed to be fixed. To this end, we first subtract the estimated baseline from \mathbf{z} to form the residual $\mathbf{r} = \mathbf{z} - \mathbf{B}\hat{\mathbf{x}}_a$ and then, compute a constrained LS estimate to obtain \mathbf{h} . To improve the estimation accuracy, we constrain the template to be localized in time (i.e., the actual template is nonzero for $L \ll N$ samples) and non-negative. Enforcing these constraints significantly improves template estimation, compared to existing blind deconvolution methods for neural data, e.g., [6]. In particular, we seek to solve the following constrained LS problem:

$$\text{(C-LS)} \quad \begin{cases} \text{minimize} & \|\mathbf{r} - \mathbf{F}^* \mathbf{\Lambda} \mathbf{F} \hat{\mathbf{x}}_a\|_2 \\ \mathbf{\Lambda} = \text{diag}(\boldsymbol{\lambda}) & \\ \text{subject to} & [\mathbf{F}^* \boldsymbol{\lambda}]_{\Omega} \geq 0, [\mathbf{F}^* \boldsymbol{\lambda}]_{\Omega^c} = 0, \end{cases}$$

where $\boldsymbol{\lambda}$ is the frequency-domain representation of the template $\mathbf{h} = \mathbf{F}^* \boldsymbol{\lambda}$. The constraints ensure that \mathbf{h} is non-negative in the set $\Omega = \{1, \dots, L\}$ and zero in its complement Ω^c .

To efficiently solve C-LS, we reformulate the problem in the time-domain by rewriting its objective function as

$$\begin{aligned} \|\mathbf{r} - \mathbf{F} \mathbf{\Lambda} \mathbf{F}^* \hat{\mathbf{x}}_a\|_2 &= \|\mathbf{F}^* \mathbf{r} - \mathbf{\Lambda} \mathbf{F}^* \hat{\mathbf{x}}_a\|_2 \\ &= \|\mathbf{F}^* \mathbf{r} - \text{diag}(\mathbf{F}^* \hat{\mathbf{x}}_a) \boldsymbol{\lambda}\|_2, \end{aligned} \quad (3)$$

where (3) exchanges the diagonal matrix $\mathbf{\Lambda} = \text{diag}(\boldsymbol{\lambda})$ with the vector $\mathbf{F}^* \hat{\mathbf{x}}_a$. By using the orthonormality property $\mathbf{F} \mathbf{F}^* = \mathbf{I}$ of the DFT matrix, we can rewrite the C-LS problem as

$$\text{(C-LS}^*) \quad \begin{cases} \text{minimize} & \|\mathbf{F}^* \mathbf{r} - \text{diag}(\mathbf{F}^* \hat{\mathbf{x}}_a) \mathbf{F} \mathbf{h}\|_2 \\ \mathbf{h} \in \mathbb{R}^N & \\ \text{subject to} & [\mathbf{h}]_{\Omega} \geq 0, [\mathbf{h}]_{\Omega^c} = 0. \end{cases}$$

This problem can be solved efficiently using an accelerated first-order method derived from the FISTA framework [11], which directly delivers an improved estimate of the template \mathbf{h} .

IV. RESULTS AND DISCUSSION

We now demonstrate the efficacy of the proposed NERDS method for synthetic and real-world data.

¹Depending on the template \mathbf{h} , matrix \mathbf{A} does in many cases not satisfy the sufficient conditions for stable recovery in [8], [9]. Nevertheless, as it will be shown in Section IV, signal separation via BP-SEP succeeds in practice.

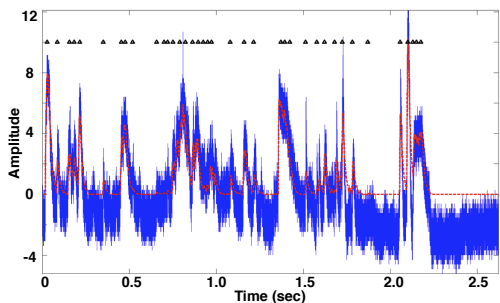


Fig. 2. Reconstruction of neuronal signal (dotted) from whole cell electrophysiology (solid). Detected spikes are shown at the top of the traces.

A. Event Detection Performance

In Fig. 2, we show an example of the proposed method for real data obtained from whole-cell current clamp recordings of differentiated human neural progenitor cells that were cultured for 17 days after removing basic fibroblast growth factor from the culture media. In addition to obtaining an estimate that is qualitatively similar to the neural signal of interest, the number of events recovered seems to agree with a careful manual analysis of the observations.

We provide results using synthetically generated calcium imaging data in Fig. 3.² These synthetic results demonstrate the advantages of BP-SEP-based recovery over BPDN, which does not account for baseline drift in the data. Moreover, we show that setting $\delta = 0$ in BP-SEP performs very similar to an oracle-based (optimal) choice of δ . The reason for this behavior is due to the observation that large part of the noise \mathbf{n} is absorbed into the coefficients of \mathbf{x}_b , thus effectively denoising the sparse vector \mathbf{x}_a . This *noise absorption phenomenon* effectively denoises the signal of interest $\mathbf{y}_a = \mathbf{A}\mathbf{x}_a$, without the need of finding an optimum parameter δ .

Our results on both real and synthetic data provide preliminary evidence that the proposed NERDS method provides a robust method for event detection that requires minimal manual intervention. In particular, NERDS only requires an estimate of the maximum length of the template L (used for template estimation), an appropriate window size Δ , and a threshold τ (both used for peak detection).

B. Template Estimation Performance

In Fig. 4, we show a manually labeled PSP event in the dataset described in the previous experiment, with our initial template estimate (following an exponential function) and the final template learned via the NERDS method. The resulting learned template exhibits a number of qualitatively similar properties to that of the original data. Concretely, we observe very similar rise and decay times, with an additional plateau in the learned template that resembles the decay properties of the true template. This preliminary result suggests that our template estimation method could provide new insights into the template waveforms that generate real neural events, rather than simply fitting the data to a pre-specified model.

²We generate signals of length $N = 4096$ with 10 exponentially decaying events and a random baseline component that is sparse in the DCT basis as in Fig. 1. We then vary the variance of the additive noise vector \mathbf{n} .

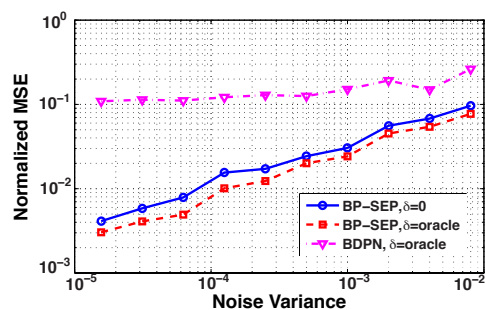


Fig. 3. Mean-squared error (MSE) comparison of standard BPDN recovery with BP-SEP for synthetic data. BP-SEP significantly outperforms BPDN. For BP-SEP, an oracle based δ performs only slightly worse than for $\delta = 0$.

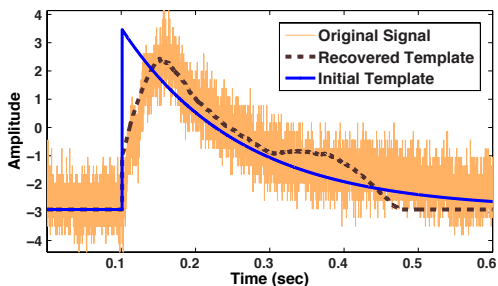


Fig. 4. Template estimation from electrophysiology Data: initial template (solid) and learned template (dash) overlaid on a manually labeled PSP. The NERDS method delivers accurate estimates of the template waveform \mathbf{h} .

REFERENCES

- [1] I. Stevenson and K. Kording, “How advances in neural recording affect data analysis,” *Nature Neurosci.*, vol. 14, pp. 139–142, Jan. 2011.
- [2] J. Clements and J. Bekkers, “Detection of spontaneous synaptic events with an optimally scaled template,” *Biophys J.*, vol. 73, no. 1, pp. 220–229, July 1997.
- [3] A. J. Pernía-Andrade, S. P. Goswami, Y. Stickler, U. Fröbe, A. Schlögl, and P. Jonas, “A deconvolution-based method with high sensitivity and temporal resolution for detection of spontaneous synaptic currents in vitro and in vivo,” *Biophys J.*, vol. 103, no. 7, pp. 1429–1439, Oct. 2012.
- [4] E. L. Dyer, M. F. Duarte, D. H. Johnson, and R. G. Baraniuk, “Recovering spikes from noisy neuronal calcium signals via structured sparse approximation,” in *Proc. 9th Intl. Conf. on Latent Variable Analysis and Signal Separation*, Sep. 2010, pp. 604–611.
- [5] J. T. Vogelstein, B. O. Watson, A. M. Packer, R. Yuste, B. Jedynek, and L. Paninski, “Spike inference from calcium imaging using sequential Monte Carlo methods,” *Biophys J.*, vol. 97, no. 2, pp. 636–655, July 2009.
- [6] C. Ekanadham, D. Tranchina, and E. Simoncelli, “A blind sparse deconvolution method for neural spike identification,” *Proc. Adv. in Neural Processing Systems (NIPS)*, vol. 24, pp. 1440–1448, Dec. 2011.
- [7] S. S. Chen, D. L. Donoho, and M. A. Saunders, “Atomic decomposition by basis pursuit,” *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, July 1998.
- [8] C. Studer, P. Kuppinger, G. Pope, and H. Bölcskei, “Recovery of sparsely corrupted signals,” *IEEE Trans. Inform. Theory*, vol. 58, no. 5, pp. 3115–3130, May 2012.
- [9] C. Studer and R. G. Baraniuk, “Recovery guarantees for restoration and separation of approximately sparse signals,” in *Proc. 49th Annual Allerton Conf. on Commun., Control, and Computing (Allerton)*, Sep. 2011, pp. 736–743.
- [10] C. Hegde, M. F. Duarte, V. Cevher *et al.*, “Compressive sensing recovery of spike trains using a structured sparsity model,” in *Proc. of Sig. Process. with Adaptive Sparse Structured Representations (SPARS)*, Mar. 2009.
- [11] A. Beck and M. Teboulle, “A fast iterative shrinkage-thresholding algorithm for linear inverse problems,” *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, Jan. 2009.