Fast spike train inference from calcium imaging

Joshua T. Vogelstein
Department of Neuroscience
Johns Hopkuns University
Baltimore, MD 21205
joshuav@jhu.edu

Baktash Babadi
Department of Neuroscience
Columbia University
New York City, NY, 10027
bb2280@columbia.edu

Brendon O. Watson
Department of Neuroscience
Columbia University
New York City, NY, 10027
bow4@columbia.edu

Rafael Yuste
Department of Neuroscience
Columbia University
New York City, NY, 10027
rmy5@columbia.edu

Liam Paninski
Department of Statistics
Columbia University
New York City, NY, 10027
liam@stat.columbia.edu

Abstract

Experiments often yield measurements of variables that are naturally constrained to be nonnegative. In such scenarios, it may be desirable to filter (or deconvolve) the observations to find the most likely trajectory of the nonnegative variable, given the noisy observations. Here, we develop a computationally-efficient optimal filter for a certain subset of nonnegatively constrained deconvolutions. Specifically, for any nonnegative variable that is filtered by a matrix linear differential equation and observed with independent, log-concave noise, we can infer the optimal nonnegative trajectory via straightforward interior-point methods in $O(T)$ (linear) time, as opposed to more standard approaches requiring $O(T^3)$ time (where T is the total number of time steps). The key is to make use of the tridiagonal structure of the Hessian of the log-posterior here, which allows us to perform each Newton iteration in linear time. We apply this filter to an important problem in neuroscience: inferring a spike train from noisy calcium fluorescence observations. We demonstrate the filter's improved performance on simulated and real data. In conclusion, we propose that this filter is readily applicable for a number of real-time applications, including spike inference from simultaneously-observed large neural populations.

1 Introduction

Experiments often yield measurements of variables known to be nonnegative, due to physical constraints. Examples arise in a wide variety of fields, including both audio and image signal processing. Therefore, determining the most likely trajectory of the nonnegative variable, given the observations, requires a filter that performs a nonnegative deconvolution. Furthermore, by imposing such a nonnegativity constraint on the solution, these constraints also regularize the resulting inference [1, 2, 3], by combating against "ringing" overfitting effects. Nonnegative deconvolution is closely related to another set of problems, called nonnegative matrix factorization, which decomposes a nonnegative matrix into the product of two nonnegative matrices [4].

Unfortunately, imposing nonnegative constraints on deconvolution or matrix factorization remains difficult. Although a number of methods have recently been introduced [5, 1, 2, 6], they all scale polynomially with the number of time steps. Here, we show that in some important special cases, we can solve these problems in linear time. The major restriction we place on the filter is that it is the solution of a matrix linear ordinary differential equation.

One important example comes from neuroscience. Calcium-sensitive fluorescent indicators are becoming an increasingly popular tool for visualizing neural spiking activity, both in vivo and in vitro [7]. Spikes cause a sharp rise in intracellular calcium concentration, $[\mathrm{Ca}^{2+}]$, which is reported by a concurrent rise in fluorescence activity [7]. As spikes are nonnegative entities, and $[\mathrm{Ca}^{2+}]$ and spikes are related by a linear matrix differential equation, this data is indeed a special case of the more general class of problems described above. We therefore develop an optimal and efficient algorithm for inferring spikes from noisy fluorescence observations, which we cast in the language of a nonnegative deconvolution problem.

The remainder of this paper is organized as follows. In Section 2 we describe a highly efficient and optimal nonnegative filter for solving the kinds of problems described above, by making use of two tools. First, we use an interior-point technique for dealing with the nonnegativity constraint, in which we iteratively solve a series of related log-concave problems. Second, when the likelihood to be maximized adheres to our constraints, we can use an efficient algorithm to solve each iteration in $O(T)$ time, where T is the total number of time steps. In Section 3, we first use simulations to compare this filter with a few other possible filters: the optimal linear (i.e., Wiener) filter, and a fast version of an algorithm referred to as projection pursuit regression (PPR), adapted to our model of interest. Then, we apply these fast filters to an example fluorescence time-series recorded from a live neuron *in vitro*, and compare with an optimal nonlinear particle filter [8]. Finally, in Section 4, we discuss some applications and extensions.

2 Methods

As mentioned above, using calcium-sensitive fluorescent indicators is becoming increasingly popular, as it enables simultaneously observing the activity of many (e.g., up to 500) neurons. Unfortunately, the image quality is typically relatively poor. To maximize the utility of this data, it would therefore be desirable to have an optimal filter, that would provide a spike train given only fluorescence measurements. Recent work towards inferring spike trains from such data have made significant advances [9, 10, 11, 12], but none have utilized the fact that spikes are nonnegative. We therefore develop such a filter, in the larger context of nonnegative filter theory.

Constructing the optimal nonnegative filter We assume the following model relating spikes, n, baseline subtracted intracellular calcium concentration, denoted by C, and fluorescence measurements, F:

$$F_t = C_t + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0, \sigma^2) \tag{1}$$

$$C_t - C_{t-1} = -\frac{\Delta}{\tau} C_{t-1} + n_t \tag{2}$$

$$n_t \sim Poisson(\lambda_t \Delta) \tag{3}$$

where ε_t is a Gaussian random variable with zero mean and σ^2 variance, τ is the time constant for calcium decay, Δ is the time step size (usually taken to be the duration of an image, i.e., $\Delta = 1/(\text{frame rate})$), and spikes are Poisson random variables with rate $\lambda_t \Delta$ (the Δ ensures that the rate does not change with temporal discretization). We emphasize that although this model is relatively simple, it has been shown to be a reasonable first-order approximation [7]. Thus, we are focusing on this relatively simple model here for clarity, and note that these assumptions may be significantly generalized, as will be discussed below.[1] Fig. 1 depicts an exemplary simulation using (1)–(3). Given such a model, our goal is to find the maximum *a posteriori* (MAP) spike train, i.e., the most likely spike train, $\boldsymbol{n} = n_1, \ldots, n_T$, where T is the final time step, given the fluorescence measurements, $\boldsymbol{F}$. We note here that spikes and calcium are related to one another via a simple linear transformation, namely, $n_t = f(C_t) = C_t - aC_{t-1}$, where $a = 1 - \Delta/\tau$. With this in mind, we may write our objective function entire in terms of C_t:

[1] Importantly, there is no parameter setting the jump size, as amplifying it by a factor would be equivalent to attenuating $\boldsymbol{n}$ by the same factor, so it would not be identifiable. Furthermore, F_t is in arbitrary units, so there is a free scale and shift term that we repressed for notational clarity. As such, neither the scale of $[\mathrm{Ca}^{2+}]$, nor the baseline, is identifiable in this model.

$$\widehat{\boldsymbol{C}}_{MAP} = \underset{f(C_t)\geq 0}{\operatorname{argmax}}\, p(\boldsymbol{C}|\boldsymbol{F}) = \underset{f(C_t)\geq 0}{\operatorname{argmax}}\, p(\boldsymbol{F}|\boldsymbol{C})p\big(f(\boldsymbol{C})\big) \tag{4a}$$

$$= \underset{f(C_t)\geq 0}{\operatorname{argmax}} \prod_{t=1}^{T} P_{\boldsymbol{\theta}}(F_t|C_{1:t})p\big(f(C_t)\big) = \underset{f(C_t)\geq 0}{\operatorname{argmax}} \sum_{t=1}^{T} \log p(F_t|C_{1:t}) + \log p\big(f(C_t)\big) \tag{4b}$$

$$\approx \underset{f(C_t)\geq 0}{\operatorname{argmax}} \sum_{t=1}^{T} -\frac{1}{2\sigma^2}(F_t - C_t)^2 - \lambda_t \Delta f(C_t) \tag{4c}$$

where (4b) follows from our model, and (4c) follows by approximating the non-convex objective in (4b) with the closest convex relaxation, as is standard in the penalized regression, deconvolution [13], and compressed sensing literature [14].

While this convex relaxation makes the problem tractable, the "sharp" threshold imposed by the nonnegativity constraint prohibits the use of standard gradient ascent techniques [15]. We therefore take an "interior-point" (or "barrier") approach, in which we drop the sharp threshold, and add a barrier term, which must approach $-\infty$ as n_t approaches zero (e.g., $-\log n_t$) [15]. By iteratively reducing the weight of the barrier term, $\eta > 0$, we are guaranteed to converge to the correct solution [15]. Thus, our goal is to efficiently solve:

$$\widehat{\boldsymbol{C}}_{\eta} = \underset{f(n_t)\forall t}{\operatorname{argmin}} \sum_{t=1}^{T} \frac{1}{2\sigma^2}(F_t - C_t)^2 + \lambda_t \Delta(C_t - aC_{t-1}) - \eta \log(C_t - aC_{t-1}) \tag{5}$$

Now, note that (2) may be written as a linear matrix equation:

$$\boldsymbol{n} = \begin{bmatrix} n_1 \\ n_2 \\ \vdots \\ n_T \end{bmatrix} = \begin{bmatrix} 1 & -\boldsymbol{a} & 0 & \cdots & \cdots \\ 0 & 1 & -\boldsymbol{a} & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 1 & -\boldsymbol{a} \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_T \end{bmatrix} = \boldsymbol{M}\boldsymbol{C} \tag{6}$$

where $\boldsymbol{M} \in \mathbb{R}^{T\times T}$ is a bidiagonal matrix. The matrix notation of (6) emphasizes that at any time t, n_t is a linear function of only C_t and C_{t-1}, which will be important for evaluating our objective function efficiently. Given the notation in (6), we can rewrite our objection function, (5), in vector notation as well:

$$\widehat{\boldsymbol{C}}_{\eta} = \underset{\boldsymbol{M}\boldsymbol{C}\geq \boldsymbol{0}}{\operatorname{argmin}} \frac{1}{2\sigma^2} \|\boldsymbol{F} - \boldsymbol{C}\|_2^2 + \boldsymbol{\lambda}\Delta\boldsymbol{M}\boldsymbol{C} - \eta \log \boldsymbol{M}\boldsymbol{C} \tag{7}$$

As (7) is concave, we may use any descent technique to find $\widehat{\boldsymbol{C}}_{\eta}$. We elect to use the Newton-Raphson approach with backtracking linesearches, i.e., update $\widehat{\boldsymbol{C}}_{\eta}$ by adding to it the solution to $\boldsymbol{H}\boldsymbol{C} = \boldsymbol{g}$, weighted by $0 < s \leq 1$, where $\boldsymbol{H}$ and $\boldsymbol{g}$ are the Hessian and gradient of the argument in (7). The step size, s, ensures that the posterior converges, by enforcing an increase in the objective with each step.

Typically, implementing Newton-Raphson requires inverting the Hessian, a computation consuming $O(T^3)$ time. Instead, because $\boldsymbol{M}$ is bidiagonal, the Hessian is tridiagonal, so the solution may be found in $O(T)$ time via standard banded Gaussian elimination techniques; the resulting fast algorithm for solving the optimization problem in (4) is the main result of this paper. In detail, we may implement this filter by iteratively solving $\boldsymbol{H}\boldsymbol{C} = \boldsymbol{g}$ (and choosing the step size) efficiently for a particular η, until the likelihood no longer increases. This process is repeated while reducing η until η approaches zero, at which time $\widehat{\boldsymbol{C}}_{\eta}$ will have converged to (4c). Fig. 1 shows how this method performs on an example fluorescence signal simulated according to (1)–(3). Note that even though the fluorescence signal is very noisy, our filter accurately infers the correct spike times.

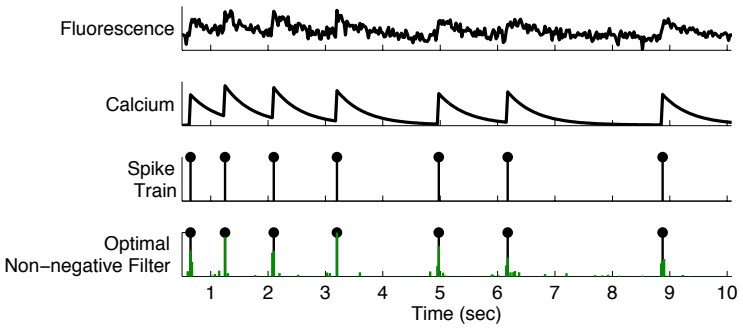

Figure 1: Simulation demonstrating our neuron model and inference. Note that our filter accurately infers the unobserved spike train, given only the noisy fluorescence measurements. Top panel: Simulated fluorescence. Second panel: Simulated intracellular calcium concentration. Third panel: Simulated spike train. Fourth panel: Output of optimal nonnegative filter (green) superimposed on simulated spike train (black). Parameters: $\lambda = 10$, $\Delta = 0.025$ msec, $\tau = 0.5$ sec, $\sigma = 1$.

Generalization of the optimal nonnegative filter The above filter design assumed a very simple model relating spikes to $[\mathrm{Ca}^{2+}]$, and $[\mathrm{Ca}^{2+}]$ to fluorescence. Both (1) and (2) may be generalized in a straightforward manner. In fact, our model may be thought of as a special case of the more general state-space framework:

$$\boldsymbol{C}_t = \boldsymbol{D}\boldsymbol{C}_{t-1} + \mathbf{1}n_t \tag{8a}$$
$$\boldsymbol{F}_t = \boldsymbol{A}\boldsymbol{C}_t + \boldsymbol{b} + \boldsymbol{\varepsilon}_t \tag{8b}$$

which we can solve in linear time for any log-concave distribution of n_t and $\boldsymbol{\varepsilon}_t$. For our particular scenario of interest, i.e., that of inferring spike trains from calcium sensitive fluorescence observations, these generalizations facilitate considering a model with much richer dynamics. For instance, a more accurate model relating spikes to $[\mathrm{Ca}^{2+}]$ would consider myriad calcium extrusion mechanisms, buffering, etc. In such a situation, we would represent $[\mathrm{Ca}^{2+}]$ as a vector, $\boldsymbol{C}_t \in \mathbb{R}^{N\times 1}$, where each element of $\boldsymbol{C}_t$ would correspond to a different spike dependent calcium process. The differential operator, $\boldsymbol{D} \in \mathbb{R}^{N\times N}$, accounts for the relative strength of each of these mechanisms, and their time constants. The spikes, n_t, are multiplied by a column vector of ones, $\mathbf{1} \in \mathbb{R}^{N\times 1}$, because the magnitude of the effect of a spike on each element in $\boldsymbol{C}_t$ is scaled by $\boldsymbol{A}$.[2] It should be clear that as in (2), spikes are related to calcium by a simple linear transformation. In the multidimensional scenario, however, $\boldsymbol{D}$ — a matrix — would replace a — a scalar — so $\boldsymbol{M}$ becomes block-bidiagonal, making the Hessian block tridiagonal. Nonetheless, Gaussian elimination may still solve $\boldsymbol{H}\boldsymbol{C} = \boldsymbol{g}$ in $O(T)$, so our filter may be applied to this scenario as well.

Another natural extension of this work would be to let F_t also be multidimensional. In (1), although we assumed that we had access to a monodimensional fluorescent magnitude at each time, the raw data is actually a multidimensional movie. These images, however, are necessarily blurred by the point-spread-function of the camera. Thus, we could replace (1) with (8b), where $\boldsymbol{A}$ is the point-spread function of the camera, and $\boldsymbol{b}$ sets the relative baseline intensity per pixel. Furthermore, the noise, $\boldsymbol{\varepsilon}_t$, could have any log-concave distribution, and these results would still hold. Given these generalizations, this filter may be applied to a large class of problems.

Learning the parameters for the optimal nonnegative filter All the above approaches assume the parameters governing our model, $\boldsymbol{\theta} = \{a, \sigma, \boldsymbol{\lambda}\}$, are known. In general, however, these parameters may be estimated from the data. To find the maximum likelihood estimator for the parameters, $\widehat{\boldsymbol{\theta}}$, we must integrate over the unknown variable, $\boldsymbol{n}$. However, integrating over all possible spike trains is typically approximated by Monte Carlo approaches, which is relatively slow. Thus, one often approximates:

[2] For the same reasons that there is no parameter scaling the spikes in the monodimensional case

$$\widehat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \int p(\boldsymbol{F}|\boldsymbol{n},\boldsymbol{\theta})p(\boldsymbol{n}|\boldsymbol{\theta})d\boldsymbol{n} \approx \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, p(\boldsymbol{F}|\widehat{\boldsymbol{n}},\boldsymbol{\theta})p(\widehat{\boldsymbol{n}}|\boldsymbol{\theta}) \tag{9}$$

where $\widehat{\boldsymbol{n}}$ is the output from the optimal nonnegative filter, i.e., the MAP estimate of $\boldsymbol{n}$. The approximation in (9) is good whenever the likelihood is very peaky, meaning that most of the mass is around the MAP sequence.[3] In particular, for state-space models, one often approximates the integral in (9) with the Viterbi path, i.e., the MAP path of the hidden states [16]. Finding the Viterbi path is often far easier than solving the integral in (9), as in the case here. The likelihood function is separable into three log-concave problems, one for each parameter. Because the noise is Gaussian, a may be estimated using standard constrained quadratic programming:

$$\widehat{a} = \underset{a>0}{\operatorname{argmax}} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{F_t - a\widehat{C}_{t-1} - \widehat{n}_t}{\sigma}\right)^2\right\} = \underset{a>0}{\operatorname{argmin}} \sum_{t=1}^{T}(F_t - a\widehat{C}_{t-1} - \widehat{n}_t)^2 \tag{10a}$$

$$= \underset{a>0}{\operatorname{argmin}} \left\|\widehat{\boldsymbol{Y}} + a\widehat{\boldsymbol{C}}\right\|_2^2 = \underset{a>0}{\operatorname{argmin}}(\widehat{\boldsymbol{C}}'\widehat{\boldsymbol{C}}a^2 + 2\widehat{\boldsymbol{Y}}'\widehat{\boldsymbol{C}}a) \tag{10b}$$

where $\widehat{\boldsymbol{Y}} = F_t - \widehat{n}_t$. Similarly, because ε_t is Gaussian, we compute $\widehat{\sigma}$ analytically using $\widehat{\sigma}^2 = (-\frac{1}{2}\widehat{\boldsymbol{C}}'\widehat{\boldsymbol{C}}\widehat{a}^2 + \widehat{\boldsymbol{Y}}'\widehat{\boldsymbol{C}}\widehat{a})/(T\Delta)$. The parameter for the prior term is given by:

$$\widehat{\boldsymbol{\lambda}} = \underset{\boldsymbol{\lambda}}{\operatorname{argmax}}\, p(\widehat{\boldsymbol{n}}|\boldsymbol{\lambda}) = \underset{\boldsymbol{\lambda}}{\operatorname{argmax}} \log p(\widehat{\boldsymbol{n}}|\boldsymbol{\lambda}) = \frac{1}{\widehat{n}_t\Delta} \tag{11}$$

which follows from assuming that $p(\boldsymbol{n}|\boldsymbol{\lambda})$ is exponential in (4c). We have found empirically that despite the approximation in (4c), upon initializing the parameters with a reasonable guess, the algorithm tends to converge to parameters that are reasonably close to the true parameters, resulting in an accurate spike reconstruction. Importantly, estimating these parameters typically requires only a very short sequence of observations, e.g., several seconds of data including about 5–10 spikes (not shown).

Projection Pursuit Regression Projection pursuit regression (PPR) is another technique one could use to infer a spike train from fluorescence measurements. PPR is different from the optimal nonnegative filter in a few ways. First, it solves a related, but slightly different problem: instead of finding the MAP estimate of the spike train, PPR finds the maximum likelihood estimate, i.e., the spike train that makes the fluorescence measurements most likely. Second, PPR constrains the solution to have only nonnegative integers. Therefore, $\boldsymbol{n}_{PPR} = \operatorname{argmax}_{n_t \in \mathbb{N}_0} p(\boldsymbol{F}|\boldsymbol{n})$, where $\mathbb{N}_0$ is the set of nonnegative integers.. Third, because of this constraint, PPR requires an additional parameter, w, that sets the size of the calcium transient caused by a single an action potential. This forces us to substitute (2) with $C_t = aC_{t-1} + wn_t$. Given these differences, to find $\boldsymbol{n}_{PPR}$, PPR proceeds iteratively, adding a spike with each iteration, as long as doing so reduces the residual square error (i.e., the sum of the squared difference between the inferred $\boldsymbol{C}$ and $\boldsymbol{F}$). One obtains the time of the next inferred spike by finding the maximum of the convolution of the current residual square error with the calcium kernel, $we^{-t\Delta/\tau}$. In general, this procedure takes $O(T \log T)$ per iteration, as the convolution is computed in the Fourier domain. However, the recursive state-space representation in (8) implies that this convolution requires just $O(T)$ time here. Henceforth, we therefore refer to this approach as fast PPR (or fPPR). This approach may generalized to the multidimensional cases, as in the previous filter. However, unlike the previous filter, the likelihood function is not concave (recall that PPR is a greedy optimization method), so we are no longer guaranteed to converge to the optimal solution.

[3]The approximation in (9) may be considered a first-order Laplace approximation

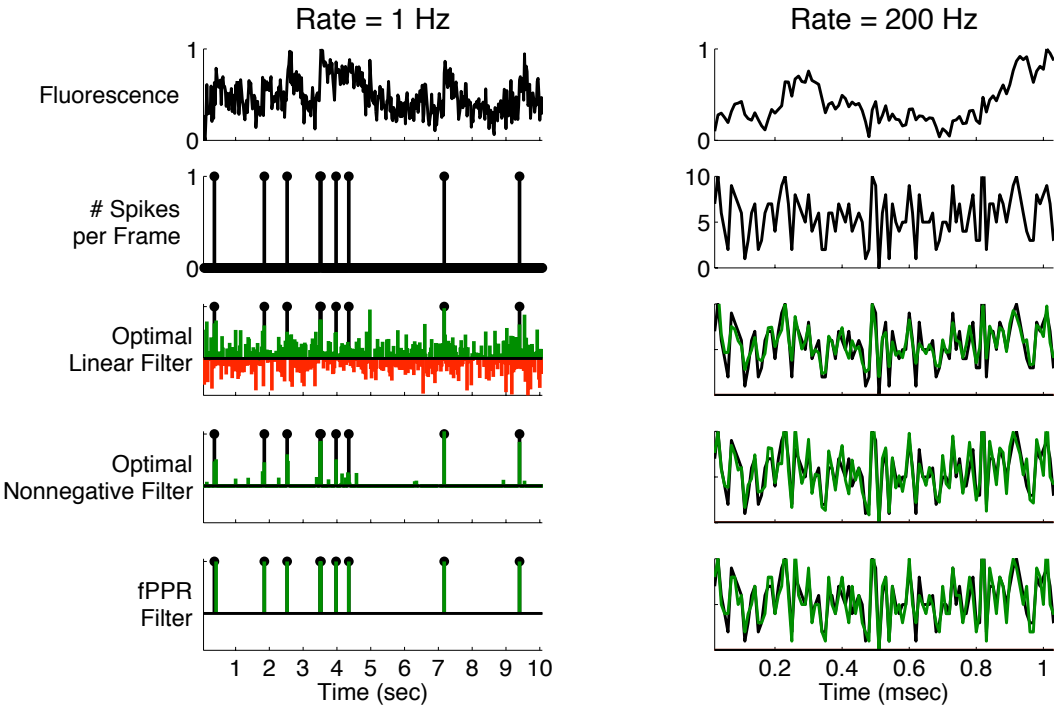

Figure 2: Comparison of various linear filters for a simulated Poisson neuron spiking with a rate of 1 Hz (left panels) and 200 Hz (right panels). The main result is that the two filters proposed outperform the optimal linear (i.e., Wiener) filter. Top panels: Fluorescence measurements. Second panels: Number of spikes per frame. Third panels: Optimal linear (i.e., Wiener) filter output given the above fluorescence signal. The Wiener filter does not impose a nonnegativity constraint, thus the inferred spike train can either be positive (green) or negative (red). Fourth panels: Same as third panels, but for the optimal nonnegative filter. Note the absence of negative spikes. Fifth panels: Same as third panels, but using the fast Projection Pursuit Regression (fPPR) filter. Parameters: $\Delta = 0.025$ sec, $\tau = 0.5$ sec, $\lambda = 1/(\text{firing rate})$, left $\sigma = 0.4$, right $\sigma = 1$.

3 Results

To evaluate the efficacy of our filters, we compare their results to that of the optimal linear (i.e., Wiener) filter [17]. The Wiener filter differs in construction from our filters in a few ways. First, it imposes no constraint on on the spike train. Second, it is optimal upon assuming that the prior distribution of n_t is Gaussian. In our model, spiking was Poisson, (3), and the nonnegativity of n_t in the sparse-spiking regime makes the Gaussian model assumed by the Wiener filter inaccurate (Fig. 2, left). However, when spike rates are fast — e.g., on average, several spikes per image frame — a Poisson distribution is better approximated by a Gaussian distribution. Furthermore, at high firing rates, the mean of the Gaussian would be relatively far from zero, and the variance proportional, so the probability of sampling a negative number would be relatively small, obviating the need for the nonnegativity constraint. Thus, one might expect the Wiener filter to perform as well as our nonnegative filter (and fPPR) when the observed neuron has a high firing rate (and relatively low imaging rate). Furthermore, given that the calcium kernel is exponential, the Wiener filter, like the filters we developed here, only requires $O(T)$ time, as opposed to the typical $O(T \log T)$.

Fig. 2 depicts a comparison between the filters developed above and the Wiener filter for two different scenarios: a slow firing rate simulation (left panels) and a fast firing rate simulation (right panels). When action potentials are sparse, the two filters we propose above outperform the Wiener filter. Moreover, at high firing rates, all three filters perform approximately equally well.

While in simulations, all the above algorithms perform reasonably well, the true test is how well they perform given data from live cells. We simultaneously recorded both electrophysiologically

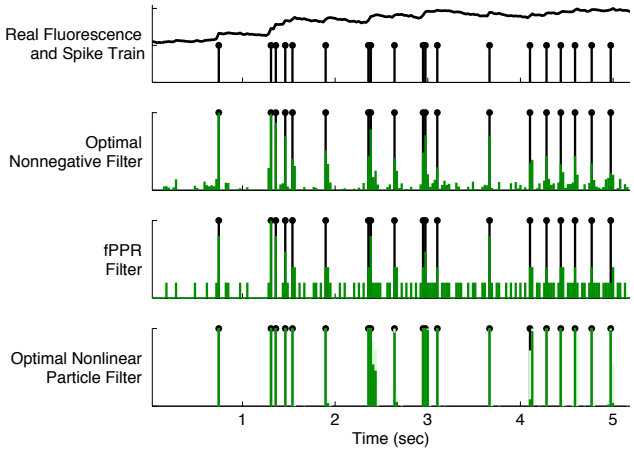

Figure 3: Inferring a spike train given fluorescence measurements from a live *in vitro* neuron. Simultaneous recording of a saturating fluorescence signal (black line in top panel), and its associated spike train (black impluses in all panels). Note how the three filters handle saturation differently. The optimal nonnegative filter's signal-to-noise ratio degrades as saturation increases, but does not suffer a catastrophic failure (green in second panel). fPPR suffers more than the optimal nonlinear particle filter when the signal saturates, due to the hard threshold (green in third panel). The optimal nonlinear particle filter [8], which explicitly incorporates saturation, correctly identifies each spike time (green in bottom panel).

and imaged with an epifluorescent microscope, from a pyramidal neuron in a somatosensory cortical slice, as described in [18]. Fig. 3 shows an example fluorescence time-series in which the neuron spiked with a relatively low rate, but the calcium accumulated nonetheless, leading to fluorescence saturation (top panel). In practice, this kind of strong saturation is rarely observed (personal communication, anonymous), so this example is designed to test the limits of performance of our filters. In fact, the optimal nonnegative filter accurately infers every spike, even when the fluorescence is strongly saturating, and the signal-to-noise ratio is very poor (second panel). On the other hand, fPPR, which has a sharp threshold for including an additional spike, performs relatively poorly (third panel), demonstrating the dependency of this method on a good model fit. For comparison purposes, we also show the performance of an optimal nonlinear particle filter [8], which specifically incorporates a nonlinear saturation function. While the optimal nonlinear particle filter performs better in scenarios such as this one, the computational burden is increased by approximately 100-fold relative to the fast nonnegative filter. Thus these two filters serve complementary purposes: the fast nonnegative filter is better geared to rapid online analysis of large-scale multi-neuronal data, whereas the nonlinear particle filter is better suited for offline refinement of the results from the fast nonnegative filter.

4 Conclusions

We show here that for certain nonnegative deconvolution problems, we can derive an algorithm that is both optimal and efficient. More specifically, our algorithm may be applied to any model with a nonnegative signal that is linearly filtered by a matrix linear ordinary differential equation. We apply this approach to the problem of inferring the most likely spike train given noisy calcium sensitive fluorescence observations (c.f. Fig. 1), and demonstrate, in simulations, that the optimal nonnegative filter outperforms the optimal linear (i.e., Wiener) filter in both slow and fast firing rate regimes (c.f. Fig. 2). Furthermore, when applied to data from a live cell, the optimal nonnegative filter outperforms a fast projection pursuit regression filter, which constrains the inferred spike train to be nonnegative integers (c.f. Fig. 3). On the other hand, the nonnegative filter is based on a linear observation model, and therefore suffers a loss of precision in the presence of strong saturation effects, in contrast to the optimal nonlinear particle filter (c.f. Fig. 3).

The implications of these results are severalfold. First, it seems as if there is no reason to use the Wiener filter for scenarios in which our algorithm may apply. Second, as our filter is so efficient, it may be used for many real-time processing applications. Specifically, upon simultaneously imaging a population of neurons [19, 20, 21, 11, 22], our filter may be applied essentially online. This could greatly expedite the tuning of important experimental parameters — such as laser intensity — to optimize signal-to-noise ratio for inferring spikes. Third, the parameters estimated from this filter may be used to initialize the parameters of the optimal nonlinear particle filter, which may then be used offline, to further refine the spike train inference. Future work will consider multidimensional models for this application, incorporating both more sophisticated calcium models, and spatial filtering for extracting the fluorescence signal, obviating the need for additional algorithms for image segmentation.

Acknowledgments Support for JTV was provided by NIDCD DC00109. LP is supported by an NSF CAREER award, by an Alfred P. Sloan Research Fellowship, and the McKnight Scholar Award. BOW was supported by NDS grant F30 NS051964. The authors would like to thank A. Packer for helpful discussions.

References

[1] Lee, D. D. and Seung, H. S. *Nature* **401**(6755), 788–791 Oct (1999).

[2] Lee, D. and Seung, H. *Advances in Neural Information Processing Systems* **13**, 556–562 (2001).

[3] Huys, Q. J. M., Ahrens, M. B., and Paninski, L. *J Neurophysiol* **96**(2), 872–890 Aug (2006).

[4] OGrady, P.D. and Pearlmutter, B.A. *Machine Learning for Signal Processing, 2006. Proceedings of the 2006 16th IEEE Signal Processing Society Workshop on* , 427–432 (2006).

[5] Portugal, L., Judice, J., and Vicente, L. *Mathematics of Computation* **63**(208), 625–643 (1994).

[6] Lee, H., Battle, A., Raina, R., and Ng, A. *Advances in Neural Information Processing Systems* (2006).

[7] Yasuda, R., Nimchinsky, E. A., Scheuss, V., Pologruto, T. A., Oertner, T. G., Sabatini, B. L., and Svoboda, K. *Sci STKE* **2004**(219), pl5 Feb (2004).

[8] Vogelstein, J., Watson, B., AM, P., Yuste, R., B, J., and Paninski, L. *Under Review* (2008).

[9] Smetters, D., Majewska, A., and Yuste, R. *Methods* **18**(2), 215–221 Jun (1999).

[10] Kerr, J. N. D., Greenberg, D., and Helmchen, F. *Proceedings of The National Academy Of Sciences Of The United States Of America* **102**(39), 14063–14068 Sep (2005).

[11] Yaksi, E. and Friedrich, R. W. *Nat Methods* **3**(5), 377–383 May (2006).

[12] Holekamp, T. F., Turaga, D., and Holy, T. E. *Neuron* **57**(5), 661–672 Mar (2008).

[13] Hastie, T., Tibshirani, R., and Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, (2001).

[14] Candes, E. *Proceedings of the International Congress of Mathematicians, Madrid, Spain* **3**, 1433–1452 (2006).

[15] Boyd, S. and Vandenberghe, L. *Convex Optimization*. Oxford University Press, (2004).

[16] Rabiner, L. R. *Proceedings of the IEEE* **72**(2), 257–286 February (1989).

[17] Wiener, N. *Extrapolation, interpolation, and smoothing of stationary time series: with engineering applications*. MIT Press, Cambridge, Mass., (1949).

[18] MacLean, J. N., Watson, B. O., Aaron, G. B., and Yuste, R. *Neuron* **48**(5), 811–823 Dec (2005).

[19] Ikegaya, Y., Aaron, G., Cossart, R., Aronov, D., Lampl, I., Ferster, D., and Yuste, R. *Science* **304**(5670), 559–564 Apr (2004).

[20] Niell, C. M. and Smith, S. J. *Neuron* **45**(6), 941–951 Mar (2005).

[21] Ohki, K., Chung, S., Ch'ng, Y. H., Kara, P., and Reid, R. C. *Nature* **433**(7026), 597–603 Feb (2005).

[22] Sato, T. R., Gray, N. W., Mainen, Z. F., and Svoboda, K. *PLoS Biol* **5**(7), e189 Jul (2007).