Inferring synaptic inputs given a noisy voltage trace via sequential Monte Carlo methods

Liam Paninski, Michael Vidne, Brian DePasquale, and Daniel Gil Ferreira
Department of Statistics and Center for Theoretical Neuroscience
Columbia University
http://www.stat.columbia.edu/∼liam

August 8, 2011

Abstract

We discuss methods for optimally inferring the synaptic inputs to an electrotonically compact neuron, given intracellular voltage-clamp or current-clamp recordings from the postsynaptic cell. These methods are based on sequential Monte Carlo techniques ("particle filtering"). We demonstrate, on model data, that these methods can recover the time course of excitatory and inhibitory synaptic inputs accurately on a single trial. Depending on the observation noise level, no averaging over multiple trials may be required. However, excitatory inputs are consistently inferred more accurately than inhibitory inputs at physiological resting potentials, due to the stronger driving force associated with excitatory conductances. Once these synaptic input time courses are recovered, it becomes possible to fit (via tractable convex optimization techniques) models describing the relationship between the sensory stimulus and the observed synaptic input. We develop both parametric and nonparametric expectation-maximization (EM) algorithms that consist of alternating iterations between these synaptic recovery and model estimation steps. We employ a fast, robust convex optimization-based method to effectively initialize the filter; these fast methods may be of independent interest. The proposed methods could be applied to better understand the balance between excitation and inhibition in sensory processing *in vivo*.

1 Introduction

The technique of *in vivo* intracellular recording holds the potential to shed a great deal of light on the biophysics of neural computation, and in particular on the dynamic balance between excitation and inhibition underlying sensory information processing (Borg-Graham et al., 1996; Peña, J.-L. and Konishi, M., 2000; Anderson et al., 2000; Wehr and Zador, 2003; Priebe and Ferster, 2005; Murphy and Rieke, 2006; Wang et al., 2007; Xie et al., 2007; Cafaro and Rieke, 2010).

However, extracting the time course of excitatory and inhibitory input conductances given a single observed voltage trace remains a difficult problem, due in large part to the fact that the problem is underdetermined: we would like to extract two variables (the excitatory and inhibitory conductances) at each time step given a single voltage observation per time step (Huys et al., 2006). Most previous work investigating the balance of excitation and inhibition during intact sensory processing has relied on averaging many trials' worth of voltage data.

1

For example, (Pospischil et al., 2007) fit average conductance quantities given a long observed voltage trace, while a number of other papers (e.g., (Borg-Graham et al., 1996; Wehr and Zador, 2003; Priebe and Ferster, 2005; Murphy and Rieke, 2006)) rely on voltage-clamping the cell at a number of different holding potentials and then averaging over a few trials in order to infer the average timecourse of synaptic inputs given a single repeated stimulus. Two important exceptions are (Wang et al., 2007), where the excitatory retinogeniculate input conductances are large and distinct enough to be inferred via direct thresholding techniques, and (Cafaro and Rieke, 2010), where the stimulus changed slowly enough that an alternating voltage-clamp and current sub-sampling strategy allowed for effectively simultaneous measurements of excitatory and inhibitory conductances.

However, it would clearly be desirable to develop methods to infer the input synaptic conductances on a trial-by-trial basis in the presence of rapidly-varying stimuli, without having to hold the cell at a variety of voltage-clamp potentials (or alternatively injecting a variety of offset currents in the current-clamp setting). Such a technique would allow us to study the variability of synaptic inputs on a fine time scale, without averaging this variability away, and would require fewer experimental trials; this is highly desirable, given the difficulty of intracellular experiments in intact preparations. Finally, these single-trial methods would open up the possibility of more detailed investigations of the spatiotemporal or spectrotemporal "receptive fields" of the excitatory and inhibitory input (Wang et al., 2007).

In this paper we develop methods based on a sequential Monte Carlo ("particle filtering") approach (Pitt and Shephard, 1999; Doucet et al., 2001) for inferring the synaptic inputs to an electrotonically compact neuron, given intracellular voltage-clamp or current-clamp recordings from the postsynaptic cell. These methods do not require averaging over many voltage traces in cases of sufficiently high SNR, are not limited to large, easily-distinguishable synaptic currents, and provide errorbars that explicitly acknowledge the uncertainty inherent in our estimates of these underdetermined quantities. Finally, these methods, which are based on well-defined probabilistic models of the synaptic input and voltage observation process, allow us to automatically infer and adapt to the noisiness of the voltage observations and the time-varying, stimulus-dependent presynaptic input rate. This is a major difference between the current work and deconvolution-based methods such as those proposed in a different context by (Yaksi and Friedrich, 2006). This well-defined stochastic model allows us to make direct connections with neural encoding models of generalized linear type (Pillow et al., 2005; Paninski et al., 2007), and could therefore lead to a more quantitative understanding of biophysical information processing.

2 Methods

In this section we describe our stochastic synaptic input model explicitly, review the necessary particle filtering methods, and derive an expectation-maximization (EM) algorithm for estimating the model parameters, along with a nonparametric version of the algorithm. Finally, we discuss an alternative optimization-based method for estimating the synaptic time courses.

2.1 The stochastic synaptic input model

For our model, we imagine that we are observing a membrane which is receiving synaptic input. Assume for now that the membrane has been made passive (Ohmic) by pharmacological

methods — i.e., we will neglect the influence of active (voltage-sensitive) channels — though this assumption may potentially be relaxed (Huys et al., 2006; Huys and Paninski, 2008). One reasonable model is

$$V(t + dt) = V(t) + dt \left[g_l(V_l - V(t)) + g_I(t)(V_I - V(t)) + g_E(t)(V_E - V(t)) \right] + \epsilon_t \quad (1)$$

$$g_I(t + dt) = g_I(t) - dt\frac{g_I(t)}{\tau_I} + N_I(t) \quad (2)$$

$$g_E(t + dt) = g_E(t) - dt\frac{g_E(t)}{\tau_E} + N_E(t), \quad (3)$$

where $N_I(t)$ and $N_E(t)$ denote the instantaneous inhibitory and excitatory inputs to the membrane at time t, respectively (Koch, 1999; Huys et al., 2006), and ϵ_t is white Gaussian noise of mean zero and variance σ_v^2. (For notational simplicity, we have suppressed the membrane capacitance as a free constant in eq. (1) and throughout the remainder of the paper.) The timestep dt is a simulation timestep, which is under the control of the data analyst, and may in general be distinct from the sampling interval of the voltage recording. We assume that the time constants τ_I and τ_E and reversal potentials V_l, V_I and V_E are known *a priori*, for simplicity (again, these assumptions may be relaxed (Huys et al., 2006)). Finally, note that we have assumed that the recordings here are in current-clamp mode (i.e., the voltage is not held to any fixed potential). We will discuss the important voltage-clamp scenario in section 2.3 below.

To perform optimal filtering in this case — i.e., to recover the time-courses of the excitatory and inhibitory conductances given the observed voltages — we need to specify a probability model for the synaptic inputs $N_I(t)$ and $N_E(t)$. The Poisson process is frequently used as a model for presynaptic inputs (Richardson and Gerstner, 2005). However, there may be many excitatory and inhibitory inputs to the membrane, with synaptic depression and facilitation thereore, $N_I(t)$ and $N_E(t)$ may not be well-represented in terms of a fixed discrete distribution on the integers, such as the Poisson. We have investigated two models that are more flexible. The first assumes that the inputs are drawn from an exponential distribution,

$$p[N_I(t)] \propto \exp(N_I(t)\lambda_I^{-1})$$

and

$$p[N_E(t)] \propto \exp(N_E(t)\lambda_E^{-1});$$

here λ_E (λ_I) denotes the mean of the excitatory (resp. inhibitory) input at time t. Thus, to be clear, from eqs. (2-3) we see that at each time step, the excitatory and inhibitory conductances are increased by independent random values which are exponentially distributed with mean λ_E and λ_I, respectively.

The second model assumes that the inputs are drawn from a truncated Gaussian distribution,

$$p[N_I(t)] \propto \exp\left(-\frac{(N_I(t) - \lambda_I)^2}{2\sigma_I^2}\right) 1[N_I(t) > 0]$$

and

$$p[N_E(t)] \propto \exp\left(-\frac{(N_E(t) - \lambda_E)^2}{2\sigma_E^2}\right) 1[N_E(t) > 0].$$

Here σ_E and σ_I denotes the spread of the excitatory and inhibitory inputs at time t around the peak λ_E and λ_I respectively. The indicator function $1[\cdot]$ is one when the argument is

3

true, and zero otherwise. The truncation at zero is due to the fact that conductances are nonnegative quantities. Also, note that the first model may be considered a special case of the second, in the limit that the curvature parameter (the "precision" σ^{-2}) approaches zero. As we will discuss further below, the filtering methods discussed here are relatively insensitive to the detailed shape of the input distributions; see also (Huys et al., 2006; Huys and Paninski, 2008).

It is clear from equations 1-3 that the dynamical variables $\big(V(t), g_E(t), g_I(t)\big)$ evolve together in a Markovian fashion. For convenience, we will abbreviate this Markovian vector as q_t, i.e., $q_t = \big(V(t), g_E(t), g_I(t)\big)$. Now we need to specify how the observed data are related to q_t. In the case that we are making direct intracellular recordings of the voltage by patch-clamp techniques, the observation noise is typically small, and the sampling frequency can be made high enough that we may assume the voltage is observed with very small noise at each time step t. However, it is frequently easier or more advantageous to use indirect voltage observation techniques (e.g., optical recordings via voltage-sensitve dye or second-harmonic generation (Dombeck et al., 2004; Nuriya et al., 2006; Araya et al., 2006)). In these cases, the noise is typically larger and the sampling frequency may be lower, particularly in the case of multiplexed observations at many different spatial locations. Thus we must model the observed voltage V_t^{obs} as a noise-corrupted version of the true underlying voltage V_t. It will be convenient (though not necessary) to assume that the observation noise is Gaussian:

$$V_t^{obs} \sim \mathcal{N}(V_t, \sigma_o^2), \tag{4}$$

where the mean is given by the true value of the voltage and the standard deviation, σ_o, is known.

2.2 Inferring synaptic input via particle filtering

Given the assumptions described above, the observed voltage data $V^{obs}(t)$ may be regarded as a sequence of observations from a hidden Markov process, $p(V_t^{obs}|q_t)$ (Rabiner, 1989). Standard methods now enable us to infer the conditional distributions, $p(q_t|\{V^{obs}\})$, of the voltage, $V(t)$, and the conductances, $g_E(t), g_I(t)$, given the observed data. In particular, this filtering problem can be solved via the technique of "particle filtering" (Pitt and Shephard, 1999; Doucet et al., 2001; Brockwell et al., 2004; Kelly and Lee, 2004; Godsill et al., 2004; Huys and Paninski, 2008; Ergun et al., 2007; Vogelstein et al., 2009), which is a Monte Carlo technique for recursively evaluating the "forward probabilities"

$$p(q_t|V_{0:t}^{obs}) = p(q_t|\{V_s^{obs}\}_{s \leq t}).$$

Using Bayes' rule these probabilities can be written in a recursive manner:

$$p(q_t|V_{0:t}^{obs}) = \int p(q_{t-dt}, q_t|V_{0:t}^{obs})dq_{t-dt} \propto \int p(V_t^{obs}|q_t)p(q_{t-dt}|V_{0:t-dt}^{obs})p(q_t|q_{t-dt})dq_{t-dt}. \tag{5}$$

Here the transition density $p(q_t|q_{t-dt})$ is provided by eqs. (1-3), and the observation density $p(V_t^{obs}|q_t)$ by eq. (4).

In most cases (including the synaptic model of interest here), one cannot compute these integrals analytically. Various numerical schemes have been proposed to compute these integrals approximately. The method introduced by (Pitt and Shephard, 1999) is most suitable for our model; specifically, we implemented what this paper refers to as the "perfectly adapted

auxiliary particle filter" (APF), which turns out to be surprisingly easy to compute in this case. The idea is to represent the forward probabilities $p(q_t|V_{0:t}^{obs})$ as a collection of unweighted samples ("particles") $\{q_t^i\}$, i.e.,

$$p(q_t|V_{0:t}^{obs}) \approx \frac{1}{N} \sum_{i=1}^{N} \delta(q_t - q_t^i);$$

here N denotes the number of particles, and q_t^i denotes the position of the i-th particle. At each time step, the positions of the particles in the $\{V, g_E, g_I\}$ space are updated in such a way as to approximately implement the integral in eq. (5). In particular, if we intepret the first equality in eq. (5) as a marginalization of the conditional joint density $p(q_{t-dt}, q_t|V_{0:t}^{obs})$ over the variable q_{t-dt}, then it is clear that all we need to do is to sample from $p(q_{t-dt}, q_t|V_{0:t}^{obs})$ (under the constraint that the other marginal, $p(q_{t-dt}|V_{0:t-dt}^{obs})$, is held fixed at the previously-computed approximation, $\frac{1}{N} \sum_{i=1}^{N} \delta(q_{t-dt} - q_{t-dt}^i)$), and then retain only the sampled q_t variables. The perfectly adapted APF accomplishes this by first sampling from $p(q_{t-dt}|V_t^{obs})$ and then from $p(q_t|q_{t-dt}, V_t^{obs})$. The algorithm simply alternates between these two steps:

1. Given the collection of particles $\{q_{t-dt}^i\}_{1 \leq i \leq N}$ at time $t - dt$, draw N auxiliary samples $\{r_{t-dt}^j\}_{1 \leq j \leq N}$ from

$$p(q_{t-dt}|V_t^{obs}) \propto \frac{1}{N} \sum_{i=1}^{N} \int p(q_t|q_{t-dt}^i)p(V_t^{obs}|q_t)dq_t.$$

2. For $1 \leq i \leq N$, draw the i-th new particle location from the conditional density $p(q_t|V_t^{obs}, q_{t-dt} = r_{t-dt}^i)$.

The resulting set of particles provide an unbiased approximation of the forward probability at the t-th time step, $p(q_t|V_{0:t}^{obs})$; then we can simply recurse forward until $t = T$, where T denotes the end of the experiment.

This APF implementation of the particle filter idea has several advantages in this context. First, since we want our algorithm to be robust to outliers and model misspecifications, it is important to sample the "correct" part of the state space, i.e., regions of the state space where the forward probability is non-negligible. In order to do that, it is beneficial to take the latest observation V_t^{obs} into consideration when proposing a new location for the particles at time step t. Second, while in general it may be difficult to sample from the necessary distributions $p(q_{t-dt}|V_t^{obs})$ and $p(q_t|V_t^{obs}, q_{t-dt})$, the necessary distributions turn out to be highly tractable in the context of this synaptic model, as we show now. Letting g_t denote the vector of excitatory and inhibitory conductances at time t (i.e., $g_t = (g_I(t), g_E(t))$), we have for the first APF step that

$$\int p(q_t|q_{t-dt})p(V_t^{obs}|q_t)dq_t = \int p(V_t, g_t|V_{t-dt}, g_{t-dt})p(V_t^{obs}|V_t)dV_t dg_t$$

$$\propto \int dV_t p(V_t|g_{t-dt}, V_{t-dt})p(V_t^{obs}|V_t) \int dg_t p(g_t|g_{t-dt})$$

$$= \int dV_t p(V_t|g_{t-dt}, V_{t-dt})p(V_t^{obs}|V_t) \qquad (6)$$

$$= \int dV_t \phi(V_t; E[V_t|g_{t-dt}, V_{t-dt}], \sigma_v^2)\phi(V_t^{obs}; V_t, \sigma_o^2)$$

$$= \phi(V_t^{obs}; E[V_t|g_{t-dt}, V_{t-dt}], \sigma_v^2 + \sigma_o^2),$$

5

where $\phi(x; \mu, \sigma^2)$ denotes the normal density with mean μ and variance σ^2 as a function of x, and $E[V_t|g_{t-dt}, V_{t-dt}]$ is given by taking the expectation of equation 1:

$$E[V_t|g_{t-dt}, V_{t-dt}] = V_{t-dt} + dt\left[g_l(V_l - V_{t-dt}) + g_I(t - dt)(V_I - V_{t-dt}) + g_E(t - dt)(V_E - V_{t-dt})\right].$$

Thus we find that the computation of the distribution in the first step of the APF is quite straightforward. Similarly, in the second step,

$$\begin{aligned}
p(q_t|q_{t-dt}, V_t^{obs}) &\propto p(q_t|q_{t-dt})p(V_t^{obs}|q_t) \\
&= p(g_t|g_{t-dt})p(V_t|g_{t-dt}, V_{t-dt})p(V_t^{obs}|V_t) \quad\quad (7)\\
&= p[g_I(t)|g_I(t - dt)]p[g_E(t)|g_E(t - dt)]p(V_t|g_{t-dt}, V_{t-dt})p(V_t^{obs}|V_t).
\end{aligned}$$

The key result here is that $g_I(t)$, $g_E(t)$, and V_t are conditionally independent given the past state q_t and the observation V_t^{obs}; in addition, the conditional distributions of $g_I(t)$ and $g_E(t)$ are simply equal to their prior distributions, and are therefore easy to sample from. To sample from the conditional distribution of V_t, we just write

$$\begin{aligned}
p(V_t|g_{t-dt}, V_{t-dt}, V_t^{obs}) &\propto p(V_t|g_{t-dt}, V_{t-dt})p(V_t^{obs}|V_t) \\
&= \phi(V_t; E[V_t|g_{t-dt}, V_{t-dt}], \sigma_v^2)\phi(V_t^{obs}; V_t, \sigma_o^2) \quad\quad (8)\\
&\propto \phi\left(V_t; \left(E[V_t|g_{t-dt}, V_{t-dt}]/\sigma_v^2 + V_t^{obs}/\sigma_o^2\right)W, W\right),
\end{aligned}$$

where we have used the standard formula for the product of two Gaussian functions and abbreviated the conditional variance as $W = (1/\sigma_v^2 + 1/\sigma_o^2)^{-1}$. Thus the second step of the APF here proceeds simply by drawing independent samples from $p[g_I(t)|g_I(t - dt)]$, $p[g_E(t)|g_E(t - dt)]$, and the above Gaussian conditional density for V_t. The resulting algorithm is fast and robust, and does not require any importance sampling, which is a basic (but unfortunately not very robust) step in more basic implementations of the particle filter (Doucet et al., 2001).

2.2.1 Backwards recursion

The APF method described above efficiently computes the collection of forward probabilities $p(q_t|V_{0:t}^{obs})$ for all desired values of time t. However, what we are interested in is the fully-conditioned probabilities, $p(q_t|V_{0:T}^{obs})$; i.e., we want to know the distribution of q_t, given all past and future observations $V_{0:T}^{obs}$, not only past observations up to the current time, $V_{0:t}^{obs}$). To compute this fully-conditioned distribution, we apply a standard "backwards" recursion:

$$p(q_t, q_{t+dt}|V_{0:T}^{obs}) = p(q_{t+dt}|V_{0:T}^{obs})\frac{p(q_{t+dt}|q_t)p(q_t, V_{0:t}^{obs})}{\int p(q_{t+dt}|q_t)p(q_t, V_{0:t}^{obs})dq_t}.$$

We represent the marginal $p(q_t|V_{0:T}^{obs})$ as

$$p(q_t|V_{0:T}^{obs}) = \sum_{j=1}^{N} u_t^{(j)}\delta(q_t - q_t^{(j)}),$$

where the particle locations $q_t^{(j)}$ are inherited from the forward step, but the particle weights $u_t^{(j)}$ are no longer simply equal to $1/N$, but instead have been redefined to incorporate the future observations, $V_{t+dt:T}^{obs}$. We initialize

$$u_T^{(i)} = 1/N,$$

and recurse backwards using the standard particle approximation to the pairwise probabilities

$$p(q_t, q_{t+dt}|V_{0:T}^{obs}) = \sum_{i=1}^{N}\sum_{j=1}^{N}\delta(q_{t+dt} - q_{t+dt}^{(j)})u_{t+dt}^{(j)}\frac{p(q_{t+dt}^{(j)}|q_t^{(i)})(1/N)\delta(q_t - q_t^{(i)})}{\sum_{l=1}^{N}p(q_{t+dt}^{(j)}|q_t^{(l)})(1/N)}, \qquad (9)$$

where $p(q_{t+dt}^{(j)}|q_t^{(i)})$ is again obtained from equation (7). Marginalizing over q_{t+dt} and canceling the $1/N$ factors, we obtain the formula for the updated backwards weight

$$u_t^{(i)} = \sum_{j=1}^{N}u_{t+dt}^{(j)}\frac{p(q_{t+dt}^{(j)}|q_t^{(i)})}{\sum_{l=1}^{N}p(q_{t+dt}^{(j)}|q_t^{(l)})}.$$

It is worth noting that this backwards recursion requires only the transition densities $p(q_{t+dt}^{(j)}|q_t^{(l)})$. The proposal density $p(q_t|V_t^{obs}, q_{t-dt})$ and observation density $p(V_t^{obs}|q_t)$ are only used in the forward recursion, and are no longer directly needed in the backwards step. In practice we found that the backwards sweep usually had just a small effect; thus one can perform just the faster forward sweep to obtain similar results at a lower computational cost.

2.2.2   Gaussian particle filter implementation: Rao-Blackwellization

One of the limiting factors of particle filters is the dimension of the state space (Bickel et al., 2008). The larger the dimension of the state space, the more particles are needed in order to achieve equal quality of representation of the underlying density. Thus it is typically beneficial to reduce the state dimensionality if possible. In particular, if some components of the posterior density of the state can be computed analytically, then it is a good idea to represent these components exactly, instead of approximating them via Monte Carlo methods. This basic idea can be stated more precisely as the Rao-Blackwell theorem from the theory of statistical inference (Casella and Berger, 2001), and therefore applications of this idea are often referred to as "Rao-Blackwellized" methods (Doucet et al., 2000).

In our case, equation 1 shows that the voltage $\{V_t\}$ is a linear Gaussian function given the synaptic conductances $\{g_t\}$. In addition we have assumed that the observations are linear and Gaussian given V_t. Thus, given the conductances $\{g_t\}$, our model may be viewed as a Kalman filter; i.e., the voltage variables $\{V_t\}$ may be integrated out analytically, reducing the effective state dimension from three to two. The basic idea is that we represent the forward probabilities as a sum of Gaussian functions now, instead of the sum of particles that we used before:

$$p(q_t|V_{0:t}^{obs}) \approx \frac{1}{N}\sum_{i=1}^{N}\delta(g_t - g_t^i)\phi(V_t; \mu_t^i, W_t^i),$$

for suitably chosen means μ_t^i and variances W_t^i see, e.g., (Kotecha and Djuric, 2003) for further discussion of this Gaussian particle representation.

Now to derive recursive updates for the conductance locations $\{g_t^i\}$, means $\{\mu_t^i\}$, and variances $\{W_t^i\}$, we simply repeat our derivation of the APF as before, now integrating out

7

the voltages V_t. The first step is to write down the conditional distribution over the indices i:

$$
\begin{aligned}
p(i|V_t^{obs}) &\propto p(V_t^{obs}|i) \\
&= \int\int p(V_t^{obs}|V_t)p(V_t|i)dV_{t-dt}dV_t \\
&= \int\int p(V_t^{obs}|V_t)p(V_{t-dt}|i)p(V_t|V_{t-dt},g_{t-dt}^i)dV_{t-dt}dV_t \\
&= \int dV_t p(V_t^{obs}|V_t)\int dV_{t-dt}p(V_{t-dt}|i)p(V_t|V_{t-dt},g_{t-dt}^i) \\
&= \int dV_t \phi(V_t^{obs};V_t,\sigma_o^2)\int dV_{t-dt}\phi(V_{t-dt};\mu_{t-dt}^i,W_{t-dt}^i)\phi(V_t;E[V_t|g_{t-dt}^i,V_{t-dt}],\sigma_v^2) \\
&= \int dV_t \phi(V_t^{obs};V_t,\sigma_o^2)\int dV_{t-dt}\phi(V_{t-dt};\mu_{t-dt}^i,W_{t-dt}^i)\phi(V_t;a_iV_{t-dt}+b_i,\sigma_v^2) \\
&= \int dV_t \phi(V_t^{obs};V_t,\sigma_o^2)\phi(V_t;a_i\mu_{t-dt}^i+b_i,a_i^2W_{t-dt}^i+\sigma_v^2) \\
&= \phi(V_t^{obs};a_i\mu_{t-dt}^i+b_i,a_i^2W_{t-dt}^i+\sigma_v^2+\sigma_o^2),
\end{aligned}
$$

$$(10)$$

where we have used the fact that, given g_{t-dt}^i, V_t is a linear-Gaussian function of V_{t-dt}, with linear coefficients $a^i = 1 - dt(g_l + g_I^i + g_E^i)$ and $b^i = dt(g_lV_l + g_I^iV_I + g_E^iV_E)$. Once we have computed these conditional probabilites $p(i|V_t^{obs})$, we draw N auxiliary index samples r_{t-dt}^j from this distribution, as before, to complete the first step of the Rao-Blackwellized APF.

Similarly, in the second step we marginalize out the voltages in $p(q_t|i,V_t^{obs})$, to sample the new conductance locations $\{g_t^i\}$, and update the means $\{\mu_t^i\}$ and variances $\{W_t^i\}$. We have

$$
\begin{aligned}
p(q_t|i,V_t^{obs}) &\propto \int p(V_{t-dt}|i)p(V_t|V_{t-dt},g_{t-dt}^i)p(g_t|g_{t-dt}^i)p(V_t^{obs}|V_t)dV_{t-dt} \\
&= p(g_t|g_{t-dt}^i)p(V_t^{obs}|V_t)\int p(V_{t-dt}|i)p(V_t|V_{t-dt},g_{t-dt}^i)dV_{t-dt} \\
&= p(g_t|g_{t-dt}^i)\phi(V_t^{obs};V_t,\sigma_o^2)\phi(V_t;a_i\mu_{t-dt}^i+b_i,a_i^2W_{t-dt}^i+\sigma_v^2) \\
&\propto p(g_t|g_{t-dt}^i)\phi(V_t;\mu_t^i,W_t^i),
\end{aligned}
$$

$$(11)$$

where

$$
(W_t^i)^{-1} = \left(a_i^2W_{t-dt}^i+\sigma_v^2\right)^{-1}+1/\sigma_o^2 \tag{12}
$$

and

$$
\mu_t^i = W_t^i\left[(a_i\mu_{t-dt}^i+b_i)\left(a_i^2W_{t-dt}^i+\sigma_v^2\right)^{-1}+V_t^{obs}/\sigma_o^2\right]. \tag{13}
$$

Thus we see that the step of sampling from the conductance has not changed — as before, we simply sample independently from $p[g_I(t)|g_I(t-dt)]$ and $p[g_E(t)|g_E(t-dt)]$ — and instead of sampling from the voltage V_t we simply update the means and variance μ_t^i and W_t^i.

Thus, finally, we have arrived at an algorithm which is robust, effective, simple to implement, and fairly fast, and may be summarized as follows.

Inferring synaptic input via Gaussian particle filtering

1. Begin at time $t - dt$ with N particles g^i_{t-dt} with corresponding voltage means μ^i_{t-dt} and variances, W^i_{t-dt}.

2. Evaluate the index probabilities $p(i|V^{obs}_t)$ using Eq.10.

3. Sample N auxiliary indices r^j_{t-dt} from a discrete distribution proportional to the index probabilities $p(i|V^{obs}_t)$.

4. Propagate the N conductance locations g^i_{t-dt} forward by sampling from $g^i_t \sim p(g_t|g^i_{t-dt})$ for $1 \le i \le N$.

5. Calculate the new means and variance μ^i_t and W^i_t via Eqs. (12-13).

6. Move to the next time step.

(To implement step 3 we used standard stratified sampling methods to minimize the variance of these samples; see, e.g. (Douc et al., 2005) for a full description.) Note that once the forward probabilities have been computed via this algorithm, it is straightforward to adapt the backwards sweep described in the last section, and finally to compute the inferred posterior means $E(q_t|V^{obs}_{0:T})$ and variances $Var(q_t|V^{obs}_{0:T})$ of the state variables. Code implementing this algorithm is available at http://www.stat.columbia.edu/~liam/research/code/synapse.zip.

2.3   Particle filtering for voltage-clamp data

Above we have discussed the problem of inferring synaptic input given voltage traces observed in current-clamp mode. It is straightforward to modify these methods to analyze current traces observed in voltage-clamp mode. One standard model of the observed current in this case is

$$I_{t+dt} = I_t + dt\left(-\frac{I_t}{\tau_i} + g_l(V_l - V_t) + g_I(t)(V_I - V_t) + g_E(t)(V_E - V_t)\right) + \eta_t, \qquad (14)$$

where V_t is the holding potential at time t[1], $g_I(t)$ and $g_E(t)$ evolve according to the usual equations (2) and (3), η_t is Gaussian noise, and τ_i is a current filtering time constant[2]. This model is mathematically equivalent to the current-clamp model discussed above if we replace V with I and g_l with $1/\tau_i$, and therefore, the particle filtering methods we have developed above apply without modification to the voltage clamp model as well. See Fig. 4 for an example of the particle filter applied to simulated voltage-clamp data.

2.4   Estimating the model parameters via expectation-maximization (EM)

The expectation-maximization (EM) algorithm is a standard method for optimizing the parameters of models involving unobserved data (Dempster et al., 1977), such as our conduc-

---

[1]We assume V_t is constant, or at least changes slowly enough that we may ignore capacitative effects, though these may potentially be included in the model as well.

[2]This time constant τ_i will typically be quite small. Mathematically speaking, it prevents any discontinuous jumps in the observed current, while physically, it may represent the lumped dynamics of the electrode and any non-space-clamped (electrotonically-distant) segments of the neuron.

tances $g_E(t)$ and $g_I(t)$, here. This algorithm is guaranteed to increase the likelihood of the model fits on each iteration, and therefore will find a local optimum of the likelihood[3].

It is fairly straightforward to develop an EM algorithm to infer the parameters of our model. In fact, it is worth generalizing the model here, in order to include the effects of any observed stimuli or other covariates on the synaptic input rates. (Subsequently, in section 2.4.1, we will discuss a nonparametric model which allows us to estimate arbitrary time-varying input rates λ_I and λ_E directly, in cases where no stimulus input information is available.) Let X_t denote the vector of all such covariates observed at time t. Several tractable models are of generalized linear form (McCullagh and Nelder, 1989). Recall that

$$N_E(t) = g_E(t) - a_E g_E(t - dt)$$

and

$$N_I(t) = g_I(t) - a_I g_I(t - dt)$$

denote the excitatory and inhibitory synaptic input at time t. Now, generalizing our previous exponential model (which does not include any stimulus dependence), we may model $N_I(t)$ as exponentially distributed with a mean that depends on X_t,

$$N_I(t) \sim \exp[f(X_t \vec{k}_I)],$$

for some decreasing function $f(\cdot)$, (i.e., the larger $X_t \vec{k}_I$ is, the larger $N_I(t)$ will be, on average, since $E(N_I(t)|X_t) = f(X_t \vec{k}_I)^{-1}$), and similarly for N_E. The vectors $\vec{k}_E$ and $\vec{k}_I$ here can be interpreted as "receptive fields" for the excitatory and inhibitory inputs, by analogy with receptive-field-type generalized linear models for spiking statistics (Simoncelli et al., 2004; Paninski et al., 2007). Similarly, we may generalize the truncated Gaussian model in a straightforward manner:

$$p[N_I(t)] \propto \exp\left(-\frac{(N_I(t) - X_t \vec{k}_I)^2}{2\sigma_I^2}\right) 1[N_I(t) > 0].$$

This is known as a "truncated regression" model in the statistics and econometrics literature (Olsen, 1978; Orme and Ruud, 2002).

To derive EM algorithms for these models, we need to write down the expected log-likelihood (Dempster et al., 1977)

$$E_{p(\{N_I(t), N_E(t)\}|\hat{\theta}^{i-1}, V_{0:T}^{obs})}\left(\log p(\{N_I(t), N_E(t), V_t\}|\theta)\right) \qquad (15)$$

and then maximize this function with respect to the model parameters $\theta = (\vec{k}_I, \vec{k}_E, \sigma)$. Here $\hat{\theta}^{i-1}$ is the estimate of the model parameters θ obtained in the previous EM iteration.

We begin with the case of exponentially-distributed inputs, $N_I(t)$ and $N_E(t)$. From our generalized linear model, we have

$$\log p(N_I(t)|X) = \log f(X_t \vec{k}_I) - N_I(t) f(X_t \vec{k}_I)$$

---

[3]In the case of Monte Carlo methods for computing the expectations needed in the EM algorithm, as in the particle filter employed here, the likelihood is no longer guaranteed to increase, due to random Monte Carlo error. However, given a sufficient number of samples (particles), the algorithm will still converge properly to a steady state, where the parameters "wobble" randomly around the location of the local likelihood maximum.

and similarly for $N_E(t)$. Thus, together with the Gaussian evolution noise model (equation 1) we have

$$
\begin{aligned}
&\log p(\{N_I(t), N_E(t), V_t\}|\theta) \\
&= \sum_t \log p(N_I(t)) + \sum_t \log p(N_E(t)) + \sum_t \log p(V_t|g_E(t), g_I(t), V_{t-dt}, \sigma^2) \\
&= \sum_t \left(\log f(X_t \vec{k}_I) - N_I(t) f(X_t \vec{k}_I) \right) + \sum_t \left(\log f(X_t \vec{k}_E) - N_E(t) f(X_t \vec{k}_E) \right) \qquad (16) \\
&\quad - \frac{1}{2} \sum_t \left(\log \sigma^2 dt + \frac{1}{\sigma^2 dt} (V_t - E[V_t|g_{t-dt}, V_{t-dt}])^2 \right) + const.,
\end{aligned}
$$

where the constant term does not depend on θ, and the nonnegativity constraints $N_I(t), N_E(t) \geq 0$ are implicit. Now, interchanging the order of the expectation in (15) with the sums above, we see that the M-step reduces to three separate optimizations:

$$
\vec{k}_I^{new} = \arg\max_{\vec{k}_I} \sum_t \left(\log f(X_t \vec{k}_I) - E(N_I(t)|\hat{\theta}^{(i-1)}, V_{0:T}^{obs}) f(X_t \vec{k}_I) \right)
$$

$$
\vec{k}_E^{new} = \arg\max_{\vec{k}_E} \sum_t \left(\log f(X_t \vec{k}_E) - E(N_E(t)|\hat{\theta}^{(i-1)}, V_{0:T}^{obs}) f(X_t \vec{k}_E) \right)
$$

$$
\begin{aligned}
\hat{\sigma}^{2\ new} &= \arg\max_{\sigma^2} -\frac{1}{2} \sum_t E_{p(g_I(t), g_E(t)|\theta^{i-1}, V_{0:T}^{obs})} \\
&\quad \left(\log \sigma^2 dt + \frac{1}{\sigma^2 dt} (V(t+dt) - V_t - dt\,[g_l(V_l - V_t) + g_I(t)(V_I - V_t) + g_E(t)(V_E - V_t)])^2 \right) \\
&= \frac{1}{dt(T/dt)} \sum_t E_{p(g_E(t), g_I(t)|\hat{\theta}^{(i-1)}, V_{0:T}^{obs})} \\
&\quad \left(V(t+dt) - V_t - dt\,[g_l(V_l - V_t) + g_I(t)(V_I - V_t) + g_E(t)(V_E - V_t)] \right)^2.
\end{aligned}
$$

Each of these optimization problems may be solved independently. The third may be solved analytically once we have computed the sufficient statistics $E(g_E(t)|\theta^{i-1}, V_{0:T}^{obs})$, $E(g_I(t)|\theta^{i-1}, V_{0:T}^{obs})$, $E(g_E(t)^2|\theta^{i-1}, V_{0:T}^{obs})$, $E(g_I(t)^2|\theta^{i-1}, V_{0:T}^{obs})$, $E(g_E(t)g_I(t)|\theta^{i-1}, V_{0:T}^{obs})$. Each of these may be estimated from the output of the particle filter. For example, we may estimate

$$
E(g_E(t)|\theta^{i-1}, V_{0:T}^{obs}) \approx \sum_i w^{(i)}(t) g_E^{(i)}(t),
$$

$$
E(g_E(t)g_I(t)|\theta^{i-1}, V_{0:T}^{obs}) \approx \sum_i w^{(i)}(t) g_E^{(i)}(t) g_I^{(i)}(t),
$$

and so on. Thus, the E-step consists of 1) a single forward-backward run of the particle filter-smoother to estimate the conditional mean and second moment of the synaptic input at each time t, given the full observed data $V_{0:T}^{obs}$, and the current parameter settings θ^{i-1},

and 2) a subsequent computation of the sufficient statistics via the sample average formulas above[4].

Further, note that $E(g_E(t)|\theta^{i-1}, V_{0:T}^{obs})$ and $E(N_E(t)|\theta^{i-1}, V_{0:T}^{obs})$ are linearly related: we obtain $g_E(t)$ by convolving $N_E(t)$ with an exponential of time constant τ_E (from equation (2)). By the linearity of expectation, we may similarly obtain $E(g_E(t)|\theta^{i-1}, V_{0:T}^{obs})$ via a linear convolution of $E(N_E(t)|\theta^{i-1}, V_{0:T}^{obs})$. Thus, if we have obtained the sufficient statistics for the third problem, we also have the sufficient statistics for the first two problems, namely $E(N_E(t)|\theta^{i-1}, V_{0:T}^{obs})$ and $E(N_I(t)|\theta^{i-1}, V_{0:T}^{obs})$.

While the third optimization above is analytically tractable, the first two typically are not (except in a special case discussed below). We must therefore perform these optimizations numerically. It is clear, since $N_I(t) \geq 0$, that the M-step for $\vec{k}_I$ reduces to a concave optimization problem in $\vec{k}_I$ whenever $f(\cdot)$ is a convex function, and $\log f(\cdot)$ is concave (and similarly for $\vec{k}_E$); thus no non-global local optima exist. (For example, we may take $f(\cdot) = \exp(\cdot)$.) Furthermore, it is quite straightforward to calculate the gradient and Hessian (second-derivative matrix) of these objective functions, and therefore, optimization via Newton-Raphson or conjugate gradient ascent methods (Press et al., 1992) is quite tractable.

In the simple case of no stimulus dependence terms in the model, it is worth noting that we may update the mean input parameters λ_I and λ_E analytically:

$$\hat{\lambda}_E^{new} = \frac{1}{T} \sum_t E(N_E(t)|V_{0:T}^{obs}, \hat{\theta}^{(i-1)}) = \frac{1}{T} \sum_t E\left[g_E(t+dt) - a_E g_E(t) \,\middle|\, V_{0:T}^{obs}, \hat{\theta}^{(i-1)} \right],$$

with the analogous update for λ_I. As usual in the EM algorithm, the M step is simply a weighted version of the usual maximum likelihood estimator (MLE) for the exponential family model (the MLE for the mean parameter of an exponential distribution is the sample mean).

The case of truncated Gaussian inputs, $N_I(t)$ and $N_E(t)$, follows nearly identically. We begin, as before, by writing down the joint loglikelihood:

$$\log p(\{N_I(t), N_E(t), V_t\}|\theta)$$
$$= \sum_t \log p(N_I(t)) + \sum_t \log p(N_E(t)) + \sum_t \log p(V_t|g_E(t), g_I(t), V_{t-dt}, \sigma^2)$$
$$= -\sum_t \left[\frac{(N_I(t) - X_t\vec{k}_I)^2}{2\sigma_I^2} + \log \int_0^\infty \exp\left(-\frac{(z - X_t\vec{k}_I)^2}{2\sigma_I^2} \right) dz \right]$$
$$- \sum_t \left[\frac{(N_E(t) - X_t\vec{k}_E)^2}{2\sigma_E^2} + \log \int_0^\infty \exp\left(-\frac{(z - X_t\vec{k}_E)^2}{2\sigma_E^2} \right) dz \right]$$
$$- \frac{1}{2} \sum_t \left(\log \sigma^2 dt + \frac{1}{\sigma^2 dt} \left(V_t - E[V_t|g_{t-dt}, V_{t-dt}] \right)^2 \right) + const.$$

---

[4]In the case of noisy or incomplete observations of the voltage $V(t)$, we need to compute three additional sufficient statistics, $E(V(t)|\theta^{i-1}, V_{0:T}^{obs})$, $E(V(t)^2|\theta^{i-1}, V_{0:T}^{obs})$, and $E(V_{t-dt}V(t)|\theta^{i-1}, V_{0:T}^{obs})$. These may be similarly estimated from the output of the particle filter, specifically equation (9). Finally, as noted in (Huys et al., 2006; Huys and Paninski, 2008), it is possible to estimate the additional model parameters, (g_l, V_l, V_E, V_I), via straightforward quadratic programming methods, once the sufficient statistics are in hand. However, if g_I and g_E have free offset terms it is not possible to uniquely specify the leak parameters, (g_l, V_l), unless observations are made at a wide range of voltages. If only a single voltage is observed, then there are more free parameters than data points and the model is not uniquely identifiable.

We find that the E step is unchanged (except for the fact that we use a truncated Gaussian instead of an exponential for the transition density in the forward-backward particle filter-smoother recursion). The update for the observation noise parameter σ is also unchanged, since the observation log-density is the same in the exponential or truncated Gaussian case. The only real difference is in the updates for the "receptive field" vectors, $\vec{k}_E$ and $\vec{k}_I$, and for the input variability parameters, σ_E and σ_I. Once again, it turns out that we may define these updates in terms of two independent concave optimization problems, (one optimization for $(\vec{k}_I, \sigma_I)$, and a separate optimization for $(\vec{k}_E, \sigma_E)$), though a slight reparameterization is required (Olsen, 1978; Orme and Ruud, 2002). We need to solve an optimization of the form

$$\left(\vec{k}_I^{new}, \sigma_I^{new}\right) = \arg\max_{\vec{k}_I, \sigma_I} \sum_t -E_{p\left(N_I(t)|\hat{\theta}^{(i-1)}, V_{0:T}^{obs}\right)} \left[\frac{(N_I(t) - X_t \vec{k}_I)^2}{2\sigma_I^2} + \log \int_0^\infty \exp\left(-\frac{(z - X_t \vec{k}_I)^2}{2\sigma_I^2}\right) dz \right].$$

We may rewrite the truncated Gaussian model as an exponential family in canonical form (Casella and Berger, 2001):

$$\begin{aligned}
\log p[N_I | \lambda_I, \sigma_I] &= -\frac{(N_I - \lambda_I)^2}{2\sigma_I^2} - \log \int_0^\infty \exp\left(-\frac{(z - \lambda_I^2)}{2\sigma_I^2}\right) dz, \ N_I > 0 \\
&= -\frac{N_I^2}{2\sigma_I^2} + \frac{N_I \lambda_I}{\sigma_I^2} - \frac{\lambda_I^2}{2\sigma_I^2} - \log \int_0^\infty \exp\left(-\frac{(z - \lambda_I^2)}{2\sigma_I^2}\right) dz, \ N_I > 0 \\
&= \rho_1 N_I^2 + \rho_2 N_I + G[\rho_1, \rho_2], \ N_I > 0
\end{aligned}$$

where we have reparametrized

$$\rho_1 = -\frac{1}{2\sigma_I}$$

and

$$\rho_2 = \frac{\lambda_I}{\sigma_I}.$$

The log-cumulant function $G[\rho_1, \rho_2]$ encapsulates all terms which do not depend on N_I. Now the key fact we need from the theory of exponential families (Casella and Berger, 2001) is that the log-likelihood $\log p[N_I | \lambda_I, \sigma_I]$ is concave as a function of the canonical parameters (ρ_1, ρ_2). In addition, the gradient and Hessian of the log-likelihood with respect to (ρ_1, ρ_2) may be computed using simple moment formulas. Since ρ_2 is a linear function of $\vec{k}_I$ if we define $\lambda_I = X_t \vec{k}_I$, and sums of concave functions are concave, this implies that our optimization problem $(\vec{k}_I, \sigma_I)$ may be recast as a concave optimization over a convex set, and again no non-global local maxima exist. (See e.g. (Olsen, 1978; Orme and Ruud, 2002) for alternate proofs of this fact.) The optimization for $(\vec{k}_E, \sigma_E)$, of course, may be handled similarly.

2.4.1 A nonparametric M-step.

How do we proceed in the important case that we have no prior information about the time course of the mean excitatory and inhibitory inputs $\lambda_I(t)$ and $\lambda_E(t)$? If we are willing to make a mild assumption about the smoothness of $\lambda_I(t)$ and $\lambda_E(t)$ as a function of t, then it turns out to be fairly straightforward to apply the EM approach, as follows. The standard trick is to slightly reinterpret X_t, the vector of "covariates": choose X as a matrix where each column is a temporally localized basis function. Now, as before, we can use the model $E[N_t] = f(X_t\theta)$; here θ is a vector of basis coefficients, and by choosing different values of

this parameter vector we can flexibly model a variety of input time-courses. For example, we chose X to consist of spline basis functions here, though any other smooth bump functions with compact support, and for which the linear equation $X\theta = \mathbf{1}$ can be solved, would be appropriate. The width of the basis function should be of a similar scale as the expected fluctuations in the mean presynaptic input: overly narrow basis functions might lead to over-fitting, while including too few basis functions will lead to oversmoothing. Since the matrix $X^T X$ is banded, optimizing the objective function in the M-step requires just $O(d)$ time, where d is the number of basis functions (Green and Silverman, 1994; Paninski et al., 2009). As before, we can now start the E-step with a flat prior, $\mu = f(X_t \theta^0)$, where $X_t \theta^0$ is chosen to be a constant as a function of time t, then use the particle filter to obtain the estimate $E[\{N_I(t), N_E(t)\}|\hat{\theta}^0, V_{0:T}^{obs}]$, then update the model parameters θ using the M-step discussed above, and iterate. (It is also possible to include regularizing penalties in the M-step (Green and Silverman, 1994; Paninski et al., 2009), but we have not found this to be necessary in our experiments to date.) See Figs. 6-8 for some illustrations of this idea.

2.5   Fast initialization via a constrained-optimization filter

The particle filtering methods discussed above are fairly computationally efficient, in the sense that the required computation time scales linearly with T, the number of time points over which we want to infer the presynaptic conductances $g_E(t)$ and $g_I(t)$. Nonetheless, it is still useful to employ a fast initialization scheme, to minimize the number of EM iterations required to obtain reasonable parameter estimates. A related strategy was recently discussed in (Vogelstein et al., 2010): the idea to use a fast optimization-based filter to initialize the more flexible but relatively computationally expensive particle filter estimation method. We can employ a similar approach here.

In (Huys et al., 2006) we pointed out that in the case of complete (i.e., noiseless and non-intermittent) voltage observations, computing the maximum *a posteriori* (MAP) estimate

$$\arg \max_{\{N_I(t), N_E(t)\}} p\left(\{N_I(t), N_E(t)\}|V_{0:T}^{obs}\right)$$

corresponds to the solution of a concave quadratic program, since $g_I(t)$ and $g_E(t)$ are linear functions of $N_I(t)$ and $N_E(t)$ (for example, $g_I(t)$ is given by the convolution $g_I(t) = N_I(t) * \exp(-t/\tau_I)$). Therefore, in the case of exponential inputs $N_E(t)$ and $N_I(t)$, the objective function may be written as $\log p(\{N_I(t), N_E(t)\}|V_{0:T}^{obs}) =$

$$\log p(V_{0:T}^{obs}|\{N_I(t), N_E(t)\}) + \log p(\{N_I(t), N_E(t)\}) + const.$$

$$= \log p(V_{0:T}^{obs}|\{N_I(t), N_E(t)\}) + \sum_{t=1}^{T}(-\lambda_E N_E(t)) + \sum_{t=1}^{T}(-\lambda_I N_I(t)) + const., \ \ N_E(t), N_I(t) \geq 0,$$

with $\log p(V_{0:T}^{obs}|\{N_I(t), N_E(t)\}) =$

$$-\frac{1}{2\sigma^2 dt}\sum_{t=2}^{T}\left[V_t - \left(V_{t-dt} + dt\big(g_l(V_l - V_t) + g_I(t)(V_I - V_t) + g_E(t)(V_E - V_t)\big)\right)\right]^2 + const.,$$

a quadratic function of $\{N_E(t), N_I(t)\}$. A similar result is obtained in the case of truncated Gaussian inputs $N_E(t)$ and $N_I(t)$.

14

We can solve this constrained MAP problem very efficiently using methods reviewed in (Paninski et al., 2010). The key insight is that an interior-point optimization approach (Boyd and Vandenberghe, 2004) here has the property that each Newton step requires just $O(T)$ time (instead of the $O(T^3)$ time that a typical matrix solve requires), due to the block-tridiagonal nature of the Hessian (second-derivative) matrix in this problem. See (Paninski et al., 2010) for full details. The resulting solution again scales linearly with T, but is faster than the particle filter method (by a factor proportional to the number of particles N). If $g_E(t)$ and $g_I(t)$ are sampled at 1 KHz, for example, then this optimization-based filtering can be performed in real time on a standard laptop (i.e., processing one second of $g_E(t)$ and $g_I(t)$ requires just one second of computation). Thus, we can use this procedure to quickly obtain an effective initialization (see Fig. 5 below).

This optimization approach may be extended to the case of noisy or incomplete voltage observations. The basic idea is to optimize over $p\left(\{V(t), N_I(t), N_E(t)\}|V_{0:T}^{obs}\right)$, i.e., we will now optimize over the unobserved voltage path V as well, instead of just over (N_E, N_I), as before. This joint log-posterior remains quadratic in (V, N_I, N_E) if the observations correspond to voltage V_t corrupted by Gaussian noise, but the log-posterior will not be jointly concave in general. Instead, this function is separately concave: concave in V with (N_E, N_I) held fixed, and concave in (N_E, N_I) with V held fixed. Thus, a natural approach is to alternate between these two optimizations until convergence. The optimization over the conductances (N_E, N_I) with the voltage V held fixed has been dealt with above. The optimization over V with (N_E, N_I) fixed may be solved by a single run of an unconstrained forward-backward Kalman filter-smoother, as discussed in (Ahmadian et al., 2011). In addition, this alternating solution can itself be effectively initialized with the solution to a closely-related concave problem: if we replace the true conductance-based likelihood term with the "current-based" approximation $\log p(V_{0:T}^{obs}|\{N_I(t), N_E(t)\}) \approx$

$$-\frac{1}{2\sigma^2 dt} \sum_{t=2}^{T} \left[V_t - \left(V_{t-dt} + dt\big(g_l(V_l - V_t) + g_I(t)(V_I - \bar{V}) + g_E(t)(V_E - \bar{V})\big)\right)\right]^2 + const.,$$

where $\bar{V}$ represents an "average" baseline voltage. (We could use $\bar{V} = V_l$, for example, or set $\bar{V}$ to be the mean observed voltage Y). Then the resulting optimization is quadratic and jointly concave in (V, N_I, N_E), and can be solved in $O(T)$ time via a straightforward adaptation of the interior-point method discussed above to appropriately constrain N_I, N_E, g_I, and g_E to be nonnegative. We find empirically that the solution to the full alternating problem is quite similar to this fast concave initialization. See (Ahmadian et al., 2011) for full details.

One natural question: why not always use the faster method? As we will see in the Results section below, despite the enhanced speed of the optimization-based approach, the particle filter method does have several advantages in terms of accuracy and flexibility. As emphasized above, the particle filter is designed to compute the desired conditional expectations, $E(N_E(t)|V_{0:T}^{obs})$ and $E(N_I(t)|V_{0:T}^{obs})$, directly. The MAP solution can in some cases, (particularly in the setting of low current noise σ), closely approximate this conditional expectation, as discussed in (Paninski, 2006a; Badel et al., 2005; Koyama and Paninski, 2010; Paninski et al., 2010; Cocco et al., 2009). But in high-noise settings, the conditional expectation and MAP solutions may diverge significantly (see (Paninski, 2006b; Badel et al., 2005) for further discussion in a related integrate-and-fire model). For example, if the noise is large enough, then the $\log p(\{N_I(t), N_E(t)\})$ term can dominate the loglikelihood term $\log p(V_{0:T}^{obs}|\{N_I(t), N_E(t)\})$; since the $\log p(\{N_I(t), N_E(t)\})$ term encourages the values of $N_I(t)$ and $N_E(t)$ to be small,

this can lead to an overly sparse solution for the inferred conductances, particularly for the inhibitory conductances, due to their weaker driving force (and correspondingly weaker log-likelihood term) at the rest potential (see Results section below for further discussion of this phenomenon). Just as importantly, the particle filter allows us to compute errorbars around our estimate $E(N_E(t)|V_{0:T}^{obs})$, by computing $Var(N_E(t)|V_{0:T}^{obs})$. But the quadratic programming method provides no such confidence intervals around the MAP solution (since the posterior is often highly non-Gaussian, due to the nonnegativity constraints on $(N_E(t), N_I(t))$). The particle filter methodology is also more flexible than the optimization approach, since the former does not rely on the concavity of the log-posterior $p\left(\{N_I(t), N_E(t)\}|V_{0:T}^{obs}\right)$. As discussed in (Huys and Paninski, 2008), this makes the incorporation of non-Gaussian observations and nonlinear dynamics fairly straightforward.

3   Results

In this section, we discuss a few illustrative applications of the methods described above. Figures 1 and 2 show simple example of the particle filter applied to simulated current-clamp data, with effectively noiseless and fairly noisy voltage observation data, respectively. To generate the observed voltage data V_t (top panel), we simulated equations (1-3) forward for the one second shown here. For this simulation (and the following examples) we used the model

$$E(N_E(t)) = \exp(k_E x_E(t))$$

$$E(N_I(t)) = \exp(k_I x_I(t)),$$

where $x_E(t)$ and $x_I(t)$ were known, sinusoidally-modulated (5 Hz) input signals ($x_I(t)$ was slightly delayed relative to $x_E(t)$, to model a burst of inhibition following the excitatory inputs), and the weights k_E and k_I are assumed known. Similar results were observed with the truncated-Gaussian input model (data not shown). Of course, we assume that $g_E(t), g_I(t), N_E(t)$, and $N_I(t)$ are not observed.

Applying the particle filter to this model leads to accurate recovery of the synaptic time courses $g_E(t)$ and $g_I(t)$. It is worth noting that the precision of our estimate for $g_E(t)$ is more accurate than that of $g_I(t)$. This is because the driving force for excitation is larger than for inhibition at this voltage. Thus changing $g_E(t)$ slightly will have a large effect on what the observed voltage should have been, whereas changing $g_I(t)$ will have a smaller effect, and therefore, our estimate for $g_I(t)$ is less constrained by the observed data. (See (Huys et al., 2006) for additional discussion of these effects.)

Figure 3 compares the behavior of the particle filter when we misspecify the mean of the synaptic inputs, $\lambda_{E/I}$. In the left column, the mean synaptic inputs are assumed to be one fifth the true values, and in the right column, they are five times larger than the actual values. Since our particle filter 'adjusts' the weights of the particles at $t - dt$ when the observation at time step t occurs, it is able to combine the prior and the observed data properly, and effectively distributes particles in the "correct" part of the $(g_E(t), g_I(t))$ space. This, in turn, leads to far more accurate inference of the true input time courses, $g_E(t)$ and $g_I(t)$.

In Fig. 4 we apply the particle filter to a voltage-clamp experiment, as described in section 2.3. We examined the behavior of the filter under two different holding potentials (-60 and -10 mV). For ease of comparison, we used the same conductance traces $g_E(t)$ and $g_I(t)$ in each experiment, generating two different observed current traces $I(t)$ via equation (14)
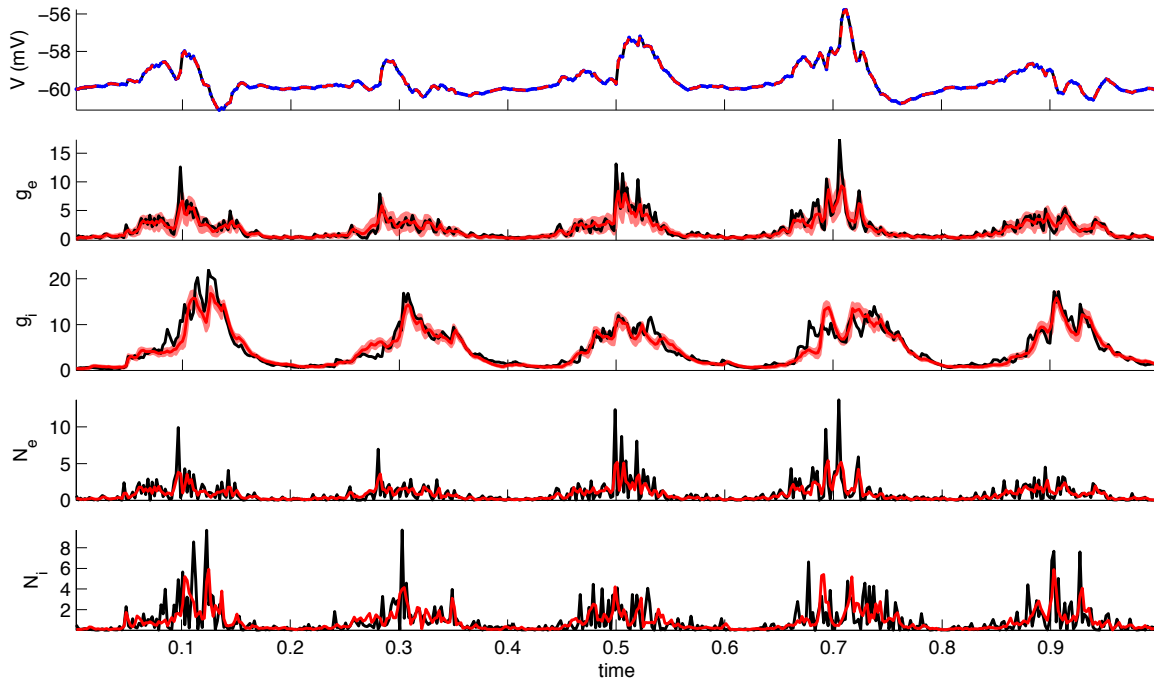
Figure 1: Estimating synaptic inputs given a single voltage trace. **Top**: True voltage in black, observed voltage data in blue, and estimated voltage in red. Since the observation noise is small they lay on top of each other. **Second and third from top**: True (black) and inferred (red) excitatory and inhibitory conductances $g_E(t)$ and $g_I(t)$, with one standard deviation. In this case, a burst of excitatory input was quickly followed by a burst of inhibition. **Bottom two panels**: True (black) and inferred (red) excitatory and inhibitory inputs $N_E(t)$ and $N_I(t)$. (Errorbars on estimated $N_E(t)$ and $N_I(t)$ are omitted for visibility.) In this simulation the time course of the mean inputs $\lambda_E(t)$ and $\lambda_I(t)$ are assumed known, but the precise timing of $N_E(t)$ and $N_I(t)$ (which are both modeled as exponential random variables given $\lambda_E(t)$ and $\lambda_I(t)$ in this and the following figures) is inferred via the particle filter technique. 100 particles were used here, with data observed and equations (1-(3)) simulated with the same timestep $dt = 2$ ms; synaptic time constants $\tau_E = 3$ ms, $\tau_I = 10$ ms; synaptic reversal potentials $V_E = 10$ mV and $V_I = -75$ mV; membrane time constant $1/g_l = 12.5$ ms, and leak potential $V_l = -60$ mV. The same parameters were used in all of the following figures unless otherwise noted. The results did not depend strongly on the parameters (data not shown).

for the two holding potentials. The true weights k_E and k_I are again assumed known. As expected given the results in Fig. 1, the particle filter recovers the true synaptic time courses $g_E(t)$ and $g_I(t)$ fairly accurately, with the accuracy for g_I improving for the more depolarized holding potential, when the inhibitory driving force is larger. Similar effects are visible in Fig. 5, which illustrates an application of the optimization-based methods discussed in section 2.5 to voltage-clamp data.

Finally, Figs. 6 and 7 address cases where we have no prior knowledge about the presynaptic stimulus timecourse. The voltage V_t (top panel) was generated using equations (1-3)
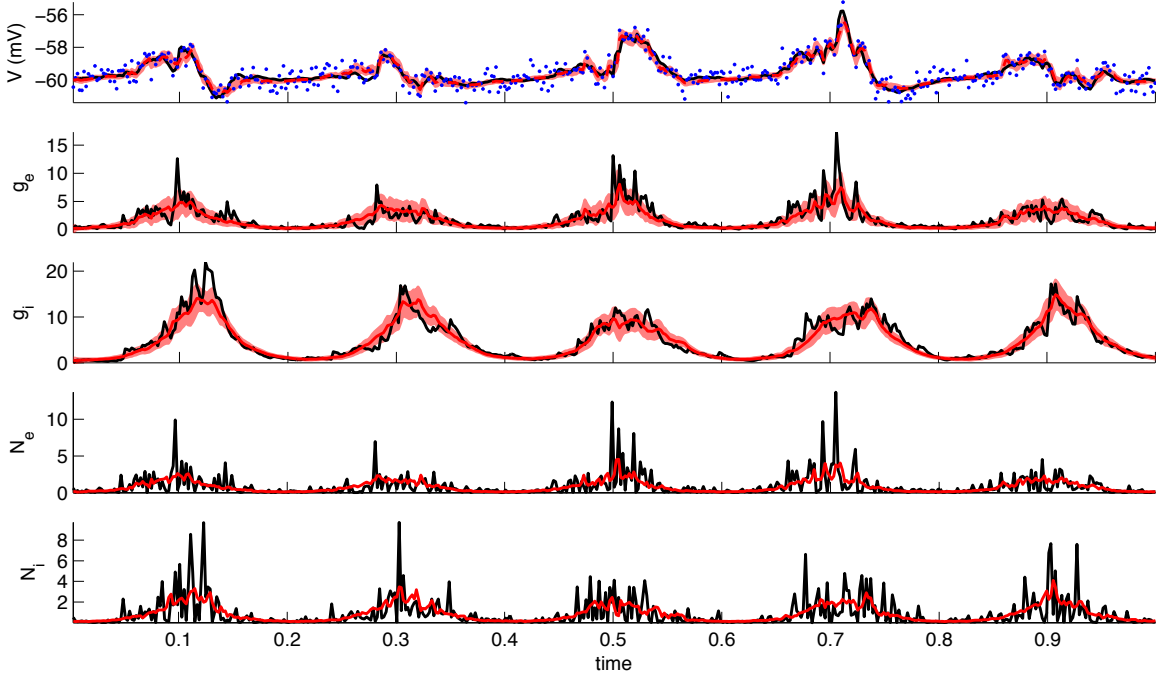
Figure 2: Estimating synaptic inputs given a voltage trace observed under noise. Conventions are as in Fig. 1, except in top panel: blue dots indicate observed noisy voltage data, while black trace indicates true (unobserved) voltage. The true voltage (black) was corrupted with white noise with zero mean and 0.44mV standard deviation. The estimates are seen to be fairly robust to noise in the observations.

forward for the 1 second shown here. However, for this simulation we used the model

$$E(N_E(t)) = \exp(\lambda_E(t))$$

$$E(N_I(t)) = \exp(\lambda_I(t)),$$

where $\lambda_E(t)$ and $\lambda_I(t)$ were the absolute value of two unobserved Ornstein-Uhlenbeck processes (these were chosen simply to generate continuous random paths λ_t; the results did not depend strongly on this choice). To infer the synaptic timecourses here, we applied the nonparametric EM method described in section 2.4.1, so that the estimated parameter $\hat{\theta}$ corresponds to the vector of weights associated with the spline basis functions that comprise the covariate matrix X. For both the parametric and nonparametric versions of the EM algorithm, we observed that on the order of 10 iterations was sufficient for convergence; for all of the examples illustrated in this paper, on the order of a second of observed data was sufficient for accurate estimates. Fig. 6 illustrates an application to a voltage trace with small observation noise, while the voltage observations in Fig. 7 are much noisier; in both cases, the method does a fairly good job inferring the synaptic time courses, although as before the inference is more accurate in the low-noise setting[5].

---

[5]It is important to make a note about the errorbars computed here. These are estimates of the posterior standard deviation $Var(g_t|V_{0:T}^{obs}, \hat{\theta})^{1/2}$, where we have conditioned on our estimate of the parameter θ. Clearly,
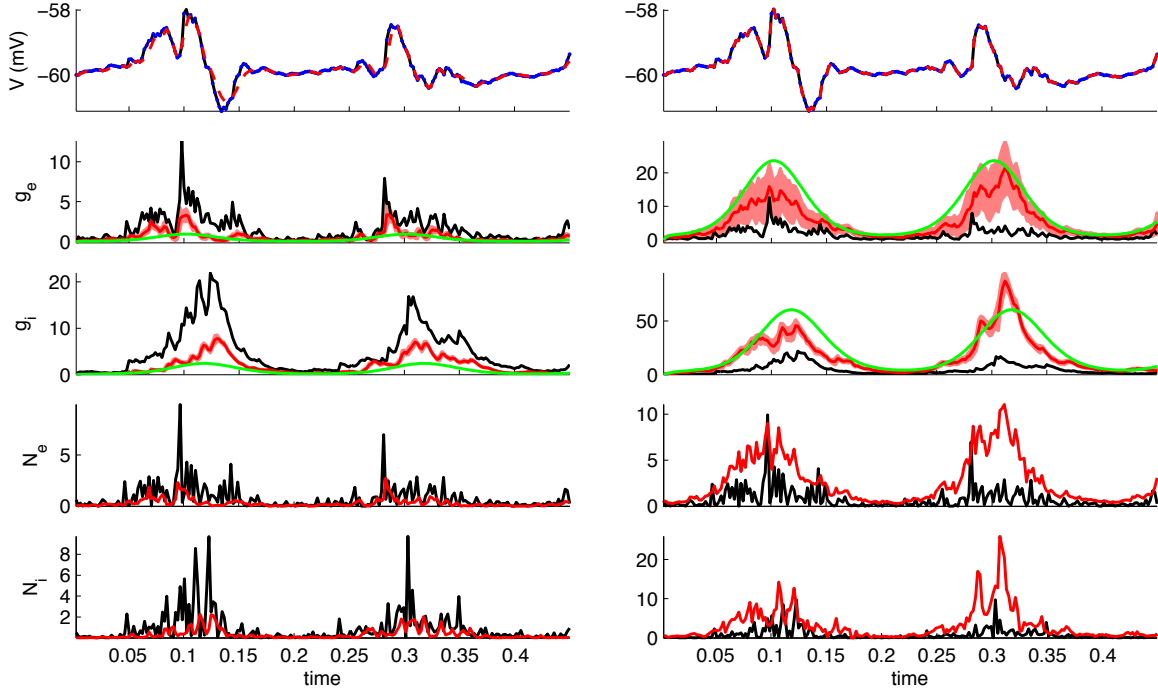
Figure 3: Comparing the robustness of the model under model misspecification. We repeated the experiment described in Fig. 1, except the particle filter was run assuming the mean synaptic inputs, $\lambda_I(t)$ and $\lambda_E(t)$, were five times as small as they in fact were in the left column and five times as large in the right column. Conventions in each column are as in Fig. 1. The misspecified expected synaptic inputs are in green. Note the particle filter effectively incorporates both the data (voltage observations) and the model prior and finds the optimal conductances and inputs according to the assumed noise parameters in the model.

Encouraged by these simulated examples, we applied our technique to a real data set provided generously by Prof. N. Sawtell. Fig. 8(top) shows a voltage trace obtained during an in vivo whole-cell recording from a granule cell in an electric fish (Sawtell, 2010). In these recordings, phasic inhibitory conductances are likely small (personal communication, N. Sawtell; though tonic inhibition is harder to rule out, this can be absorbed into the leak conductance term), facilitating the evaluation of the algorithm's performance. We applied our nonparametric algorithm with parameters chosen as follows: $V_E = 0$ mV, $V_l = -76$ mV, $1/g_l = 6$ ms, and $\tau_E = 5$ ms; the inference seemed relatively insensitive to the details of these parameters. As in the simulated examples, we initialize the algorithm with a flat mean excitatory input (using 76 spline functions in this case) and let the EM algorithm converge (about 10 iterations were again sufficient). Figure 8 shows the end result of the estimation procedure; while ground truth presynaptic recordings are not yet available in this preparation, the results of the analysis seem qualitatively reasonable. We hope to pursue further applications to real data in the future.

---

this will be an underestimate of our true posterior uncertainty, which should also incorporate our uncertainty about $\hat{\theta}$. It is possible to employ Markov chain Monte Carlo methods to incorporate this additional uncertainty about $\hat{\theta}$ (Gelman et al., 2003), but we have not yet pursued this direction.
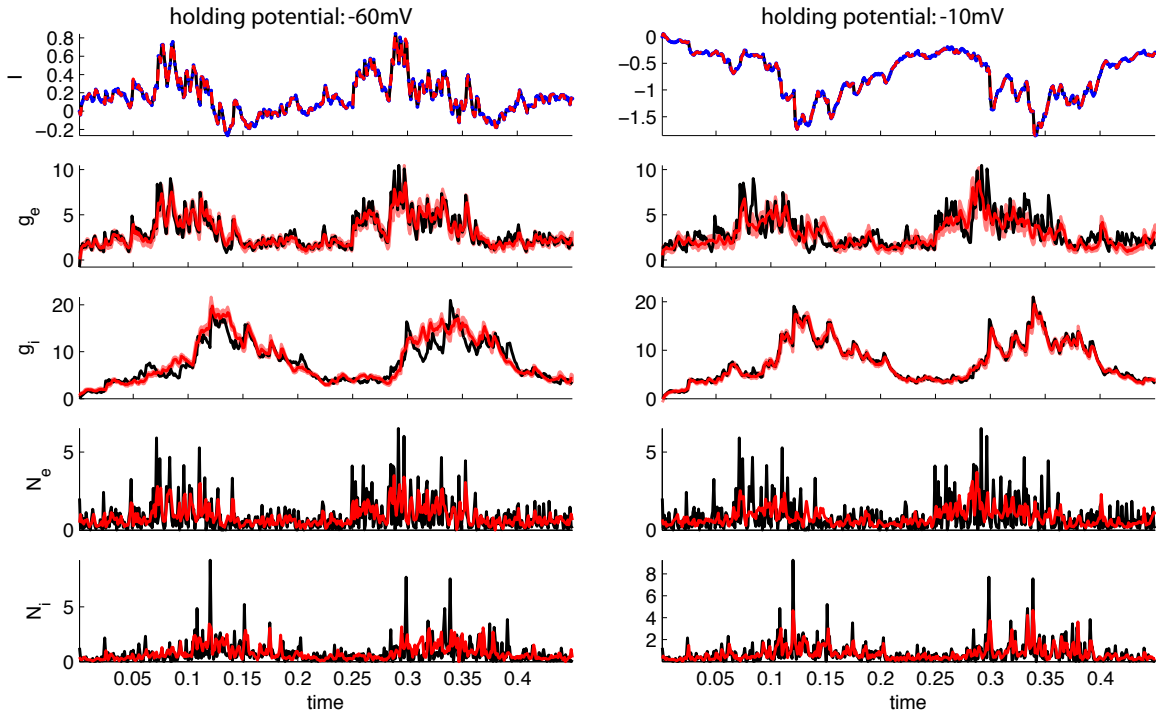
Figure 4: Estimating synaptic inputs given a noisy current trace under voltage-clamp. Two different experiments (at different holding potentials) are shown in the left and right columns. **Top**: True current in black, observed current data in blue, and estimated current in red. **Second and third panels**: True (black) and inferred (red) excitatory and inhibitory conductances $g_E(t)$ and $g_I(t)$; in this case, a burst of excitatory input was quickly followed by a burst of inhibition. **Bottom two panels**: True (black) and inferred (red) excitatory and inhibitory inputs $N_E(t)$ and $N_I(t)$. As in Fig. 1, the time course of the mean inputs $\lambda_I(t)$ and $\lambda_E(t)$ are assumed known, but the precise timing of $N_E(t)$ and $N_I(t)$ (which are both modeled as exponential random variables given $\lambda_I(t)$ and $\lambda_E(t)$) is inferred via the particle filter technique (100 particles used here). Note that, as in Fig. 1, the inference of the excitatory input is more accurate than that of the inhibitory input in the left panels (and vice versa in the right panels), due to the larger driving force of the excitatory conductances at the -60 mV holding potential.

4   Conclusion

We have introduced an robust and computationally efficient method for inferring excitatory and inhibitory synaptic inputs given a single observed noisy voltage trace. As in (Vogelstein et al., 2010), we have found that a fast, robust optimization-based filter provided a good initialization for a more computationally-intensive, expectation-maximization-based particle filter method. As expected, the inference achieved by these methods was not perfect. In particular, it is consistently more difficult to recover the details of inhibitory presynaptic input at physiological resting potentials, due to the larger driving force associated with excitatory conductances. Nonetheless, we believe that the systems and computational neuroscience community will find these methods a useful complement to the methods already available,
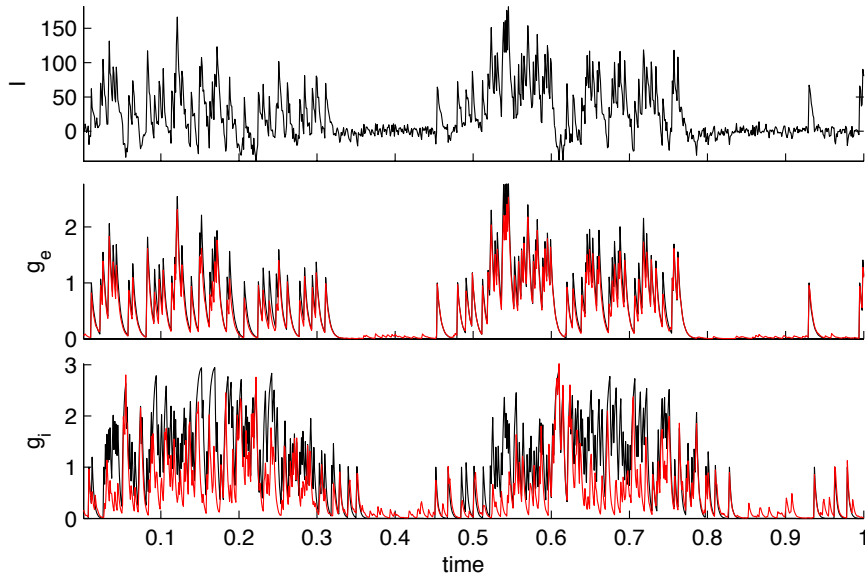
Figure 5: Using a fast optimization-based method to estimate conductances. A voltage-clamp experiment was simulated here. **Top**: observed current $I(t)$; holding potential -60 mV. **Lower panels**: true (black) and estimated (red) conductances $g_E(t)$ and $g_I(t)$. Presynaptic inputs $N_I(t)$ and $N_E(t)$ were generated via an inhomogeneous Poisson process whose rate varied sinusoidally at 2 Hz; the phase of the inhibitory sine wave was displaced by 20 ms relative to the excitatory sine carrier. This information was not made available to the filter; i.e., a constant input rate was assumed when inferring these inputs from the observed $I(t)$. Note that, as above, the excitatory inputs are estimated well, while the estimates of the inhibitory inputs are less accurate, due to the larger driving force of the excitatory conductances.

which require experimenters to hold the neuron at a variety of holding potentials and often complicate the simultaneous analysis of correlated excitatory and inhibitory input into neurons.

Several possible extensions of these methods are readily apparent. First, as discussed in (Huys et al., 2006; Huys and Paninski, 2008; Paninski et al., 2009), it is conceptually straightforward to handle multi-compartmental neural models, or to incorporate temporally-colored noise sources. These extensions may be especially useful in the context of poorly space-clamped recordings (i.e., voltage-clamp recordings in electrotonically non-compact cells). Further possible extensions include the incorporation of active membrane conductances and NMDA-gated (voltage-sensitive) synaptic channels (Huys et al., 2006; Huys and Paninski, 2008), although each of these extensions may come at a significant cost in computational expense and in the amount of data required to obtain accurate estimates of the model parameters. We could also potentially use a model in which the presynaptic inputs $N_I(t)$ and $N_E(t)$ are correlated, extending the expectation-maximization method to infer the correlation parameters. Finally, it may be necessary to incorporate a more accurate model of the filtering effects of the electrode when analyzing voltage data recorded via patch clamp or sharp electrode (Brette et al., 2007). We plan to explore the importance of these issues for the analysis of real data in future work.
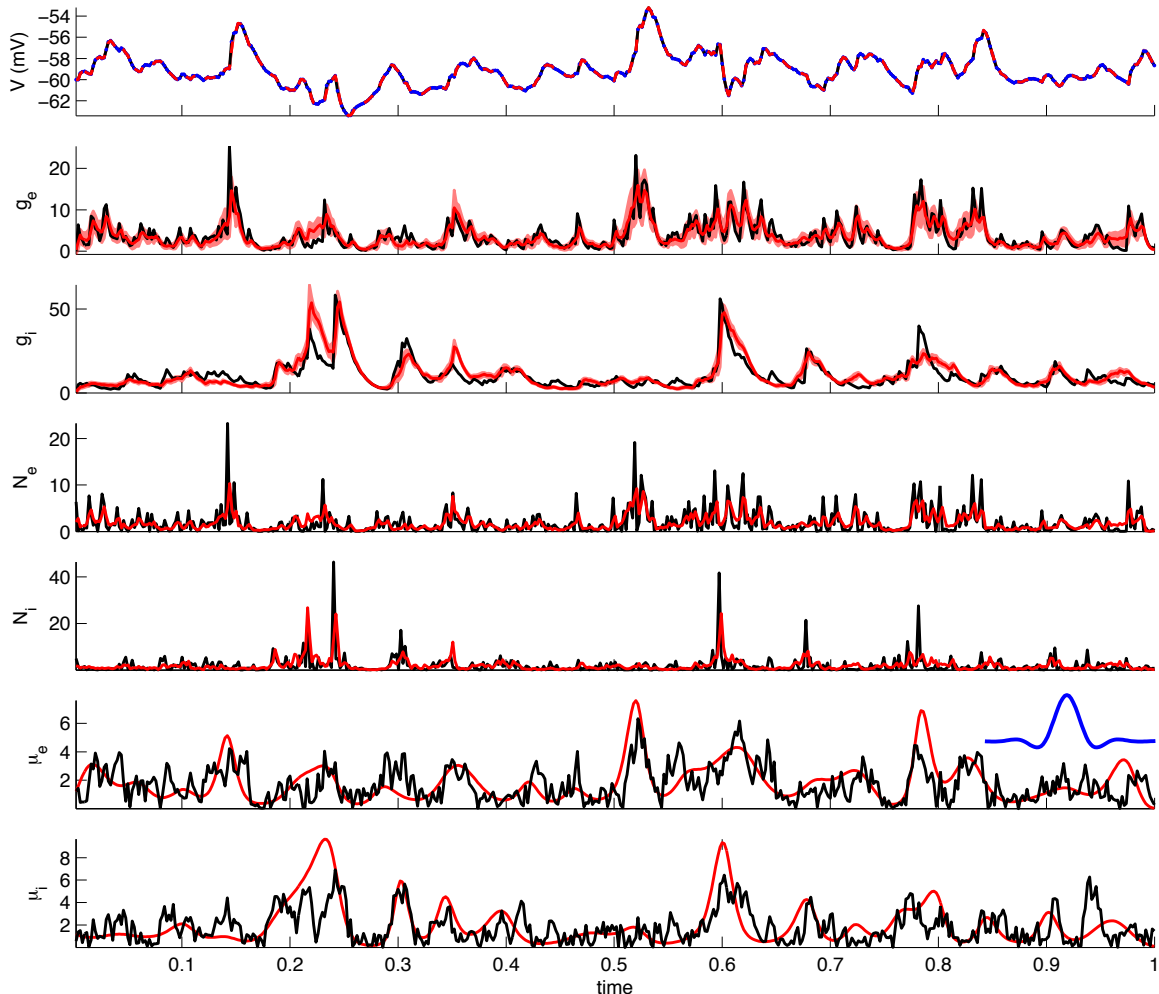
Figure 6: Estimating synaptic inputs for an arbitrary unknown mean presynaptic input, via a nonparametric EM approach. Conventions are as in Fig. 1, with the exception of the bottom two panels, which show the true (black) and estimated (red) mean presynaptic inputs $\lambda_E(t)$ and $\lambda_I(t)$; these were generated as the absolute value of random realizations of Ornstein-Uhlenbeck processes and were not known by the algorithm. $N_E(t)$ and $N_I(t)$ are both modeled as exponential random variables given $\lambda_E(t)$ and $\lambda_I(t)$. We start the estimation with a flat prior; the covariate matrix X was composed of 50 spline basis functions (an example basis function is shown in blue) and iterate the nonparametric EM algorithm (section 2.4.1)until it converges (less then 10 iterations here).
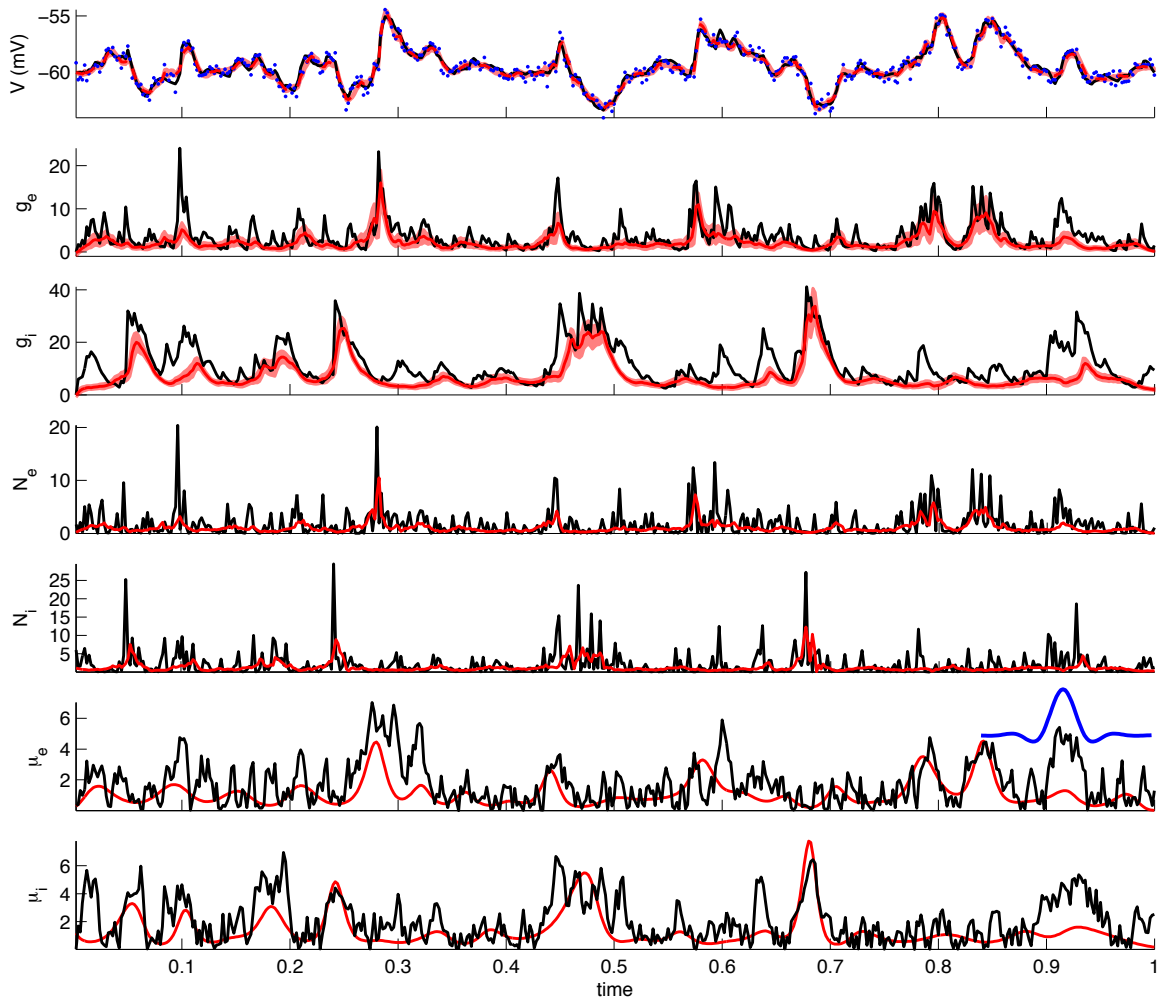
Acknowledgments

Figure 7: Estimating synaptic inputs for an arbitrary unknown mean presynaptic input given a noisy voltage trace. Conventions are as in Fig. 6. The two bottom most panels are the true (black) presynaptic stimulus and its estimation. The data is generated in the same manner as in figure 6 but the true voltage trace is corrupted with white noise of zero mean and 0.44mV standard deviation.

2010).

References

Ahmadian, Y., Packer, A., Yuste, R., and Paninski, L. (2011). Designing optimal stimuli to control neuronal spike timing. *Journal of Neurophysiology*.

Anderson, J., Carandini, M., and Ferster, D. (2000). Orientation tuning of input conductance, excitation, and inhibition in cat primary visual cortex. *Journal of Neurophysiology*, 84(2):909.
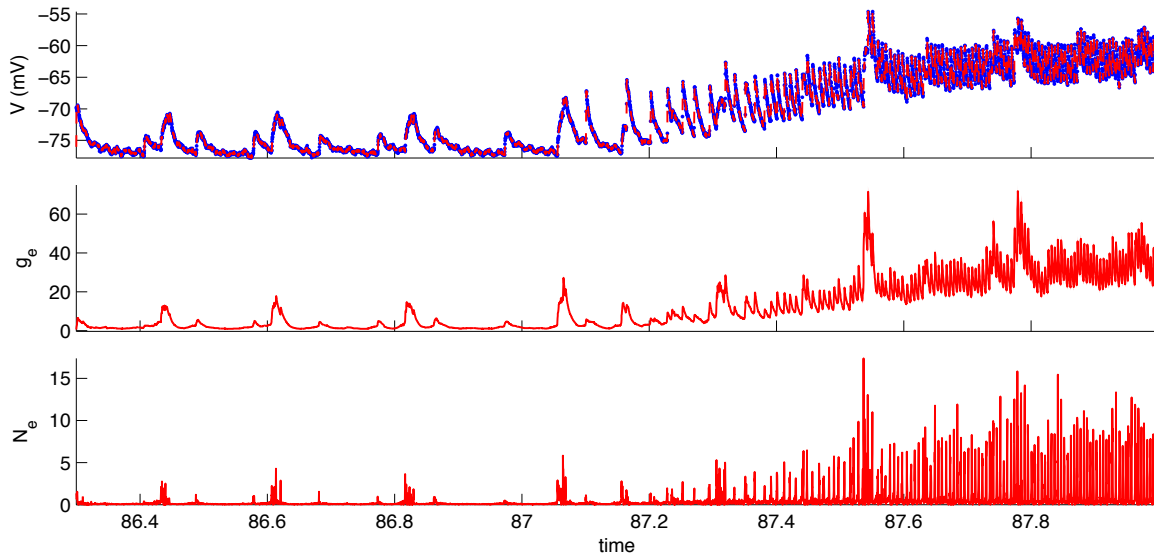
Figure 8: Estimating the excitatory conductance of an electric fish granule cell. The top panel is the recorded voltage (blue) and the estimated voltage (red) (they largely coincide since we assume low observation noise here); the 2nd panel is the estimated excitatory conductance $g_E(t)$; and the 3rd panel is the estimated synaptic input $N_E(t)$ to the cell. While comparisons to ground truth are not yet possible in this case, the inferred $g_E(t)$ and $N_E(t)$ seem entirely consistent with the observed voltage data.

Araya, R., Jiang, J., Eisenthal, K. B., and Yuste, R. (2006). The spine neck filters membrane potentials. *PNAS*, 103(47):17961–17966.

Badel, L., Richardson, M., and Gerstner, W. (2005). Dependence of the spike-triggered average voltage on membrane response properties. *Neurocomputing*, 69:1062–1065.

Bickel, P., Li, B., and Bengtsson, T. (2008). Sharp failure rates for the bootstrap particle filter in high dimensions. *IMS Collections 2008, Vol. 3, 318-329*.

Borg-Graham, L., Monier, C., and Fregnac, Y. (1996). Voltage-clamp measurement of visually-evoked conductances with whole-cell patch recordings in primary visual cortex. *Journal of Physiology-Paris*, 90(3-4):185–188.

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

Brette, R., Piwkowska, Z., Rudolph, M., Bal, T., and Destexhe, A. (2007). A nonparametric electrode model for intracellular recording. *Neurocomputing*, 70:1597–1601.

Brockwell, A., Rojas, A., and Kass, R. (2004). Recursive Bayesian decoding of motor cortical signals by particle filtering. *Journal of Neurophysiology*, 91:1899–1907.

Cafaro, J. and Rieke, F. (2010). Noise correlations improve response fidelity and stimulus encoding. *Nature*, 468(7326):964–967.

Casella, G. and Berger, R. (2001). *Statistical inference*. Duxbury Press.

Cocco, S., Leibler, S., and Monasson, R. (2009). Neuronal couplings between retinal ganglion cells inferred by efficient inverse statistical physics methods. *Proceedings of the National Academy of Sciences*, 106(33):14058–14062.

Dempster, A., Laird, N., and Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal Royal Stat. Soc., Series B*, 39:1–38.

Dombeck, D., Blanchard-Desce, M., and Webb, W. (2004). Optical recording of action potentials with second-harmonic generation microscopy. *The Journal of neuroscience*, 24(4):999.

Douc, R., Cappe, O., and Moulines, E. (2005). Comparison of resampling schemes for particle filtering. *Proc. 4th Int. Symp. Image and Signal Processing and Analyis*.

Doucet, A., De Freitas, N., and Gordon, N. (2001). *Sequential Monte Carlo methods in practice*. Springer Verlag.

Doucet, A., De Freitas, N., Murphy, K., and Russell, S. (2000). Rao-Blackwellised particle filtering for dynamic bayesian networks. In *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence*, pages 176–183. Citeseer.

Ergun, A., Barbieri, R., Eden, U., Wilson, M., and Brown, E. (2007). Construction of point process adaptive filter algorithms for neural systems using sequential Monte Carlo methods. *IEEE Transactions on Biomedical Engineering*, 54:419–428.

Gelman, A., Carlin, J., Stern, H., and Rubin, D. (2003). *Bayesian Data Analysis*. CRC Press.

Godsill, S., Doucet, A., and West, M. (2004). Monte Carlo smoothing for non-linear time series. *Journal of the American Statistical Association*, 99:156–168.

Green, P. and Silverman, B. (1994). *Nonparametric Regression and Generalized Linear Models*. CRC Press.

Huys, Q., Ahrens, M., and Paninski, L. (2006). Efficient estimation of detailed single-neuron models. *Journal of Neurophysiology*, 96:872–890.

Huys, Q. and Paninski, L. (2008). Model-based smoothing of, and parameter estimation from, noisy biophysical recordings. *PLOS Computational Biology*, Under review.

Kelly, R. and Lee, T. (2004). Decoding V1 neuronal activity using particle filtering with Volterra kernels. *Advances in Neural Information Processing Systems*, 15:1359–1366.

Koch, C. (1999). *Biophysics of Computation*. Oxford University Press.

Kotecha, J. H. and Djuric, P. M. (2003). Gaussian particle filtering. *IEEE Transactions on Signal Processing*, 51:2592–2601.

Koyama, S. and Paninski, L. (2010). Efficient computation of the maximum a posteriori path and parameter estimation in integrate-and-fire and more general state-space models. *Journal of Computational Neuroscience*, 29(1):89–105.

McCullagh, P. and Nelder, J. (1989). *Generalized linear models*. Chapman and Hall, London.

Murphy, G. and Rieke, F. (2006). Network variability limits stimulus-evoked spike timing precision in retinal ganglion cells. *Neuron*, 52:511–524.

Nuriya, M., Jiang, J., Nemet, B., Eisenthal, K., and Yuste, R. (2006). Imaging membrane potential in dendritic spines. *PNAS*, 103:786–790.

Olsen, R. J. (1978). Note on the uniqueness of the maximum likelihood estimator for the tobit model. *Econometrica*, 46:1211–1215.

Orme, C. D. and Ruud, P. A. (2002). On the uniqueness of the maximum likelihood estimator. *Economics Letters*, 75:209 – 217.

Paninski, L. (2006a). The most likely voltage path and large deviations approximations for integrate-and-fire neurons. *Journal of Computational Neuroscience*, 21:71–87.

Paninski, L. (2006b). The spike-triggered average of the integrate-and-fire cell driven by Gaussian white noise. *Neural Computation*, 18:2592–2616.

Paninski, L., Ahmadian, Y., Ferreira, D., Koyama, S., Rahnama Rad, K., Vidne, M., Vogelstein, J., and Wu, W. (2009). A new look at state-space models for neural data. *Journal of Computational Neuroscience*, pages 1–20.

Paninski, L., Ahmadian, Y., Ferreira, D., Koyama, S., Rahnama Rad, K., Vidne, M., Vogelstein, J., and Wu, W. (2010). A new look at state-space models for neural data. *Journal of Computational Neuroscience*, 29(1):107–126.

Paninski, L., Pillow, J., and Lewi, J. (2007). Statistical models for neural encoding, decoding, and optimal stimulus design. In Cisek, P., Drew, T., and Kalaska, J., editors, *Computational Neuroscience: Progress in Brain Research*. Elsevier.

Peña, J.-L. and Konishi, M. (2000). Cellular mechanisms for resolving phase ambiguity in the owl's inferior colliculus. *Proceedings of the National Academy of Sciences of the United States of America*, 97:11787–11792.

Pillow, J., Paninski, L., Uzzell, V., Simoncelli, E., and Chichilnisky, E. (2005). Prediction and decoding of retinal ganglion cell responses with a probabilistic spiking model. *Journal of Neuroscience*, 25:11003–11013.

Pitt, M. and Shephard, N. (1999). Filtering via simulation: Auxiliary particle filters. *Journal of the American Statistical Association*, 94(446):590–599.

Pospischil, M., Piwkowska, Z., Rudolph, M., Bal, T., and Destexhe, A. (2007). Calculating event-triggered average synaptic conductances from the membrane potential. *Journal of Neurophysiology*, 97:2544–2552.

Press, W., Teukolsky, S., Vetterling, W., and Flannery, B. (1992). *Numerical recipes in C*. Cambridge University Press.

Priebe, N. and Ferster, D. (2005). Direction selectivity of excitation and inhibition in simple cells of the cat primary visual cortex. *Neuron*, 45:133–145.

Rabiner, L. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77:257–286.

Richardson, M. J. E. and Gerstner, W. (2005). Synaptic shot noise and conductance fluctuations affect the membrane voltage with equal significance. *Neural Comp.*, 17(4):923–947.

Sawtell, N. (2010). Multimodal integration in granule cells as a basis for associative plasticity and sensory prediction in a cerebellum-like circuit. *Neuron*, 66(4):573–584.

Simoncelli, E., Paninski, L., Pillow, J., and Schwartz, O. (2004). Characterization of neural responses with stochastic stimuli. In *The Cognitive Neurosciences*. MIT Press, 3rd edition.

Vogelstein, J., Packer, A., Machado, T., Sippy, T., Babadi, B., Yuste, R., and Paninski, L. (2010). Fast nonnegative deconvolution for spike train inference from population calcium imaging. *Journal of Neurophysiology*, 104(6):3691.

Vogelstein, J., Watson, B., Packer, A., Jedynak, B., Yuste, R., and Paninski, L. (2009). Model-based optimal inference of spike times and calcium dynamics given noisy and intermittent calcium-fluorescence imaging. *Biophysical Journal*, 97:636–655.

Wang, X., Wei, Y., Vaingankar, V., Wang, Q., Koepsell, K., Sommer, F., and Hirsch, J. (2007). Feedforward excitation and inhibition evoke dual modes of firing in the cat's visual thalamus during naturalistic viewing. *Neuron*, 55:465–478.

Wehr, M. and Zador, A. (2003). Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex. *Nature*, 426:442–446.

Xie, R., Gittelman, J., and Pollak, G. (2007). Rethinking tuning: in vivo whole-cell recordings of the inferior colliculus in awake bats. *The Journal of Neuroscience*, 27(35):9469.

Yaksi, E. and Friedrich, R. W. (2006). Reconstruction of firing rate changes across neuronal populations by temporally deconvolved Ca^{2+} imaging. *Nature Methods*, 3(5):377–383.