

Fast Kalman filtering on quasilinear dendritic trees

Liam Paninski

Received: 16 July 2009 / Revised: 3 November 2009 / Accepted: 13 November 2009
© Springer Science + Business Media, LLC 2009

Abstract Optimal filtering of noisy voltage signals on dendritic trees is a key problem in computational cellular neuroscience. However, the state variable in this problem—the vector of voltages at every compartment—is very high-dimensional: realistic multicompartmental models often have on the order of $N = 10^4$ compartments. Standard implementations of the Kalman filter require $O(N^3)$ time and $O(N^2)$ space, and are therefore impractical. Here we take advantage of three special features of the dendritic filtering problem to construct an efficient filter: (1) dendritic dynamics are governed by a cable equation on a tree, which may be solved using sparse matrix methods in $O(N)$ time; and current methods for observing dendritic voltage (2) provide low SNR observations and (3) only image a relatively small number of compartments at a time. The idea is to approximate the Kalman equations in terms of a low-rank perturbation of the steady-state (zero-SNR) solution, which may be obtained in $O(N)$ time using methods that exploit the sparse tree structure of dendritic dynamics. The resulting methods give a very good approximation to the exact Kalman solution, but only require $O(N)$ time and space. We illustrate the method with applications to real and simulated dendritic branching structures, and describe how to extend the techniques to incorporate spatially subsampled, temporally filtered, and nonlinearly transformed observations.

Keywords State-space model · Voltage-sensitive imaging · Calcium-sensitive imaging · Markov random field

1 Introduction

The problem of understanding dendritic computation remains a key open challenge in cellular and computational neuroscience (Stuart et al. 1999; Spruston 2008; Sjostrom et al. 2008). The major difficulty is in recording physiological signals (especially voltage) with sufficient spatiotemporal resolution on dendritic trees. As pointed out in Morse et al. (2001), Wood et al. (2004), Huys et al. (2006), if we are given the full spatiotemporal voltage signal on the dendritic tree, then simple, direct statistical methods allow us to infer many biophysical quantities of interest, including passive cable parameters (e.g., the axial resistance at each point on the tree), active properties (e.g., the spatial membrane density distribution of voltage-gated channels), and in some cases even time-varying information (e.g., synaptic weights and presynaptic input conductances (Cox 2004; Paninski and Ferreira 2008; Paninski et al. 2009)). Unfortunately, multiple-electrode recordings from dendrites are quite technically challenging, and provide spatially-incomplete observations (Stuart and Sakmann 1994; Cox and Griffith 2001; Cox and Raol 2004; Bell and Craciun 2005; Petrusca et al. 2007), while high-resolution imaging techniques provide more spatially-complete observations, but with significantly lower signal-to-noise (Djurisic et al. 2004; Sacconi et al. 2006; Araya et al. 2006; Palmer and Stuart 2006; Gobel and Helmchen 2007; Vucinic and Sejnowski 2007; Canepari et al. 2007; Djurisic et al. 2008).

Action Editor: Wolfram Gerstner

Liam Paninski (✉)
Department of Statistics and Center for Theoretical
Neuroscience, Columbia University, New York, NY, USA
e-mail: liam@stat.columbia.edu
URL: <http://www.stat.columbia.edu/~liam>

One avenue for extending the reach of these currently available methods is to develop statistical techniques for optimally combining, filtering, and deconvolving these noisy signals. As argued in Huys and Paninski 2009, state-space filtering methods are attractive here, since these methods allow us to quite transparently incorporate 1) realistic, spatially-complex multicompartmental models of dendritic dynamics and 2) time-varying, heterogeneous observations (e.g., spatially-scanned multiphoton imaging data) into our filtering equations.

The problem is that the time-varying state vector in this problem—which includes, at least, the vector of voltages at every compartment—is very high-dimensional: realistic multicompartmental models often have on the order of $N \sim 10^4$ compartments. Standard implementations of state-space filter methods (e.g., the Kalman filter) require $O(N^3)$ time and $O(N^2)$ space, and are therefore impractical for applications to large dendritic trees. (Note that the analyses in Huys and Paninski 2009 were restricted to relatively low-dimensional systems.)

However, we may take advantage of three special features of the dendritic filtering problem to construct efficient filtering methods. First, dendritic dynamics are governed by a cable equation on a tree, which may be solved using symmetric sparse matrix methods in $O(N)$ time (Hines 1984). Second, current methods for imaging dendritic voltage provide low SNR observations, as discussed above. Finally, in typical experiments we record only a few image observations ($n < 100$ or so coarse pixels) at a time. Taken together, these special features allow us to approximate the Kalman equations in terms of a low-rank perturbation of the steady-state (zero-SNR) solution, which in turn may be obtained in $O(N)$ time using efficient matrix solving methods that exploit the sparse tree structure of the dynamics¹. The resulting methods provide a very good approximation to the exact Kalman solution, but only require $O(N)$ time and space. Some $O(n^3)$ operations are required; thus n is the real bottleneck for this problem, not N .

We present a derivation of this fast tree Kalman filter method in Section 2 below. We illustrate the

method with some applications to simulated, spatially-subsampled observations on large, real dendritic trees in Section 3. Finally, in Section 4 we discuss a number of key extensions of the basic method: in particular, we can incorporate spatially blurred or scanned observations; temporally filtered observations and inhomogeneous noise sources on the tree; “quasi-active” membrane dynamics (Koch 1984; Manwani and Koch 1999; Coombes et al. 2007; Kellems et al. 2009); and even in some cases nonlinear observations of the membrane state. Conclusions and directions for future research are discussed in Section 5.

2 Basic method

We begin by describing the state-space model employed here. Most of our analysis is based on a simple linear-Gaussian (Kalman) model for the noisy dynamics of the dendritic voltage and for the observed data (though we will consider more general nonlinear observations in Section 4.3 below). This linear-Gaussian model is of course a crude approximation, but serves as a reasonable starting point for understanding subthreshold (non-spiking) passive dynamics along dendritic trees; see Huys and Paninski (2009) for further background and discussion. We take the state vector here to include the vector of voltages on the full dendritic tree: if we have broken up the tree into N compartments, then our hidden state V_t has dimensionality N . (See Section 4.2 for a generalization.) For the dynamics and observation equations, we take

$$V_{t+dt} = AV_t + i_t dt + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \sigma^2 dt I) \quad (1)$$

$$y_t = B_t V_t + \eta_t, \quad \eta_t \sim \mathcal{N}(\mu_t^y, W_t). \quad (2)$$

Here A is an $N \times N$ matrix that implements the cable equation (as discussed further below), i_t is an N -vector of known, possibly time-varying currents that are injected into the compartments, and ϵ_t denotes a Gaussian noise source that perturbs the cable equation stochastically on each time step; $\mathcal{N}(\mu, C)$ denotes a Gaussian density of mean μ and covariance C . We assume for now that the covariance of ϵ_t is proportional to the identity matrix, though we may generalize this assumption somewhat; see Section 4.2 below. Note that it is fair to assume that membrane channel noise on a dendrite is spatially uncorrelated (corresponding to a diagonal covariance matrix for ϵ_t); synaptic noise, on the other hand, may not be completely spatially uncorrelated between compartments (due to branching presynaptic axons forming multiple terminals on the

¹Kellems et al. (2009) recently introduced powerful model-reduction methods for decreasing the effective dimensionality of the state vector in large dendritic voltage simulations; these methods are distinct from the techniques we introduce in this paper, but there are a number of interesting connections between these methods which we will describe below.

observed postsynaptic cell), but modeling this source of noise as uncorrelated as well seems to be a reasonable first approximation. The noise is assumed to have zero mean (with no loss of generality, due to the inclusion of the i_t term) scaled by a standard deviation σ and a \sqrt{dt} factor to ensure that the behavior of the model is insensitive to the choice of the timestep dt .

The second equation determines the observed data vector y_t . B_t is a matrix that specifies how the observations are related instantaneously to the voltage vector;² for example, in the case of a whole-cell recording with a single electrode, B_t would be an N vector all of zeros, except at the compartment whose voltage is being recorded, while in an imaging experiment B_t could encode instantaneous optical blurring and sampling transforming the N -element voltage vector V_t to the n -pixel observed vector y_t . More generally, we could combine imaging and electrode observations, simply by appending the corresponding rows of B_t . It is also worth noting that our methods are sufficiently general to allow the observation gain matrix B_t , offset mean μ_t^y , and noise covariance W_t to vary with time; in fact, even the dimensionality of y_t could vary with time (e.g., in the case of scanning optical experiments). However, neither A nor σ in the dynamics equation are allowed to vary with time.

We will begin by discussing a forward Euler discretization for the cable equation matrix A :

$$V_{t+dt}(x) = V_t(x) + dt \left(-g_x V_t(x) + i_t(x) + \sum_{w \in N(x)} a_{xw} [V_t(w) - V_t(x)] \right) + \epsilon_t(x),$$

where $V_t(x)$ denotes the voltage in compartment x at time t , g_x is the membrane conductance in compartment x , $i_t(x)$ and $\epsilon_t(x)$ represent the known current and noise current, respectively, injected into compartment x at time t , and $N(x)$ is the set of nearest neighbors of compartment x , i.e., the set of all compartments adjoining x ;³ $a_{xw} = a_{wx}$ denotes the intracellular conductance between compartments x and w . The choice of forward

Euler here is for simplicity only, to keep the derivations below as transparent as possible; we will discuss more efficient implicit implementations of the discrete-time cable equation in Section 2.1 below. The key fact about the matrix A that implements this forward Euler propagation is that it is symmetric and “tree-tridiagonal” (also known as of “Hines” form Hines (1984)): all off-diagonal elements A_{xw} are zero unless the compartments x and w are nearest neighbors on the dendritic tree.

Now the focus of this paper is the efficient implementation of the Kalman filter recursion (Durbin and Koopman 2001) for computing the forward mean

$$\mu_t = E(V_t | Y_{1:t})$$

and covariance

$$C_t = Cov(V_t | Y_{1:t}),$$

where $Y_{1:t}$ denotes all of the observed data $\{y_s\}$ up to time t . The Kalman recursions are⁴:

$$C_t = \left[(AC_{t-dt}A^T + \sigma^2 dt I)^{-1} + B^T W^{-1} B \right]^{-1}$$

$$\mu_t = C_t \left[(AC_{t-dt}A^T + \sigma^2 dt I)^{-1} (A\mu_{t-dt} + i_t dt) + B^T W^{-1} (y_t - \mu_t^y) \right].$$

We have suppressed the possible time-dependence of B and W in the observation equation for notational clarity; the extension of these equations to the general case is standard (Durbin and Koopman 2001). Note that the computation of the inverses in the recursion for C_t requires $O(N^3)$ time in general, or $O(N^2)$ time (via the Woodbury lemma) if the observation matrix B is of low rank. In either case, $O(N^2)$ space is required to store C_t .

These recursions are typically initialized with the marginal equilibrium covariance (i.e., the steady-state covariance of V_t in the absence of any observations):

$$C_0 = \lim_{t \rightarrow \infty} Cov(V_t).$$

This equilibrium covariance C_0 satisfies the discrete Lyapunov equation (Brockwell and Davis 1991)

$$AC_0A^T + \sigma^2 dt I = C_0. \tag{3}$$

²In some cases it is overly crude to assume that the observations y_t are linearly and instantaneously related to the voltage V_t ; we will discuss extensions to this basic setup in Section 4 below.

³We will assume throughout this paper that the structural anatomy of the dendritic tree (i.e., the neighborhood structure $N(x)$) is fully known.

⁴For simplicity, we focus exclusively on the forward Kalman filter recursion in this paper, to compute $p(V_t | Y_{1:t})$, but note that all of the methods discussed here can be applied to obtain the forward-backward (Kalman smoother) density $p(V_t | Y_{1:T})$, for $t < T$.

This equation can be solved explicitly here: applying the standard moving-average recursion (Brockwell and Davis 1991) for the autoregressive model V_t leads to

$$C_0 = \sum_{i=0}^{\infty} A^i cov(\epsilon)(A^T)^i \tag{4}$$

in general, and in our case, $cov(\epsilon) = \sigma^2 dt I$ and $A^T = A$, so this reduces to the explicit solution

$$C_0 = \sigma^2 dt \sum_{i=0}^{\infty} A^{2i} = \sigma^2 dt (I - A^2)^{-1}$$

whenever the dynamics matrix A is stable (all eigenvalues have absolute value less than one); for the cable equation written in forward Euler form, A may be guaranteed to be stable when the timestep dt is sufficiently small. Note that we do not represent C_0 directly, as this would require $O(N^2)$ storage and computation time. If desired, we can compute the diagonal elements of C_0 (corresponding to the marginal equilibrium variances of the voltages at each compartment x) via the junction tree algorithm (Jordan 1999; Weiss and Freeman 2001; Shental et al. 2008), which is guaranteed to require just $O(N)$ time (and space) because the potential matrix $C_0^{-1} \propto (I - A^2)$ has a generalized tree structure: $[C_0^{-1}]_{xw}$ is nonzero only if compartments x and w are nearest- or second-nearest neighbors.

With these preliminaries out of the way, we can move on to the main contribution of this paper. The basic idea is that, in low-SNR conditions, C_t should

be close to C_0 : i.e., we should be able to represent the time-varying covariance C_t as a small perturbation about the steady-state solution C_0 , in some sense. This intuition is illustrated in Fig. 1. If we compare C_t to C_0 by computing the spectrum of $C_0^{-1}C_t$, we see that only a small fraction of the eigenvalues of $C_0^{-1}C_t$ are significantly different from one. Thus $C_0^{-1}C_t$ may be approximated as a low-rank perturbation of the identity matrix, or equivalently, C_t may be approximated as a low-rank perturbation of C_0 . The effective rank of this perturbation will depend on the SNR of the observations, specifically on the scale of the observation noise W and on the dimensionality n of the observation vector y_i : the larger n is, or the smaller W is, the larger the effective rank of $I - C_0^{-1}C_t$.

Thus, more concretely, we will approximate C_t as

$$C_t \approx C_0 + U_t D_t U_t^T,$$

where $U_t D_t U_t^T$ is a low-rank matrix we will update directly. We can compute and update the perturbations U_t and D_t in $O(N)$ time, as discussed in detail below. We should note that a great deal of related work on low-rank approximations to C_t has appeared in the engineering and applied math literature (Verlaan 1998; Treebushny and Madsen 2005; Gillijns et al. 2006; Chandrasekar et al. 2008). However, we are not aware of previous low-rank approximations specialized to the case of low-SNR observations; note that, as discussed above, there is no good justification for approximating C_t itself as low-rank in this setting.

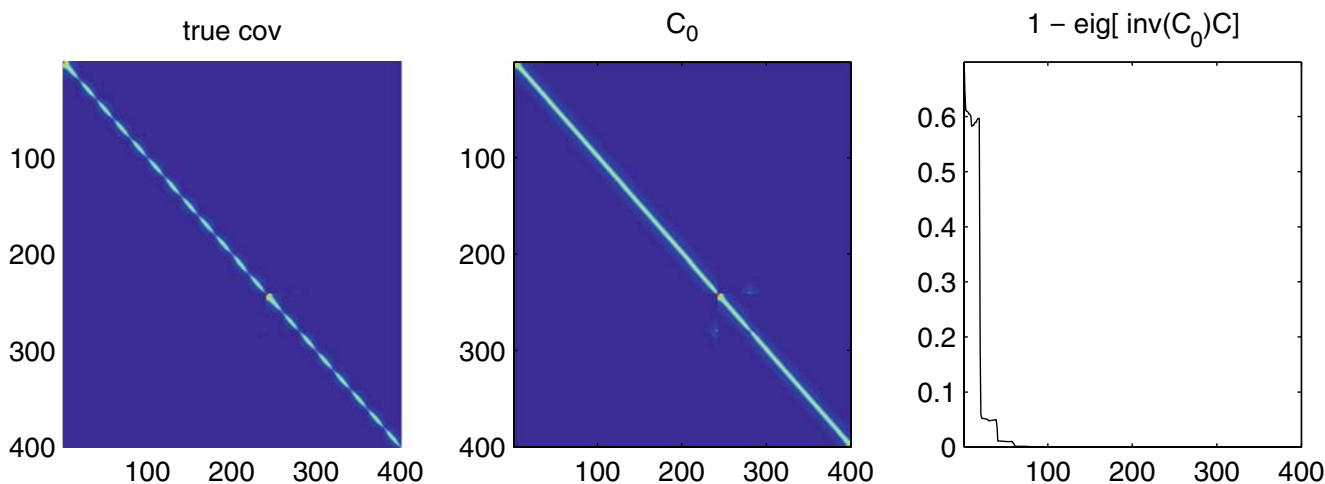


Fig. 1 C_t is fairly close to C_0 ; in particular, $I - C_0^{-1}C_t$ has low effective rank. $N = 400$. $n = 20$: each 20-th compartment is observed. Observation noise set to be about 1/2 of the marginal variance. This neuron had one branch; the branch point is visible at about compartment 230. *Left*: true C_t (t is taken to be large enough that C_t has converged to its limit, the solution of the

corresponding Riccati equation (Durbin and Koopman 2001)). *Middle*: C_0 . Both C_0 and C have been divided by $\max(C_0)$, to improve visibility. *Right*: spectrum of $I - C_0^{-1}C_t$. Note that the magnitude of the eigenvalues falls sharply after the 20-th eigenvalue, and again after the 40-th eigenvalue; an approximation of rank about 60 would seem to suffice for $I - C_0^{-1}C_t$ here

We now derive the $O(N)$ low-rank updates for μ_t , U_t , and D_t . First, write

$$\begin{aligned} & (AC_{t-dt}A^T + \sigma^2 dtI)^{-1} \\ &= (A[C_0 + U_{t-dt}D_{t-dt}U_{t-dt}^T]A^T + \sigma^2 dtI)^{-1} \\ &= (C_0 + AU_{t-dt}D_{t-dt}U_{t-dt}^T A^T)^{-1}; \end{aligned}$$

the second equality follows from Eq. (3). Now apply the Woodbury matrix lemma:

$$\begin{aligned} & (C_0 + AU_{t-dt}D_{t-dt}U_{t-dt}^T A^T)^{-1} \\ &= C_0^{-1} - C_0^{-1}AU_{t-dt} \\ &\quad \times (D_{t-dt}^{-1} + U_{t-dt}^T A^T C_0^{-1}AU_{t-dt})^{-1}U_{t-dt}^T A^T C_0^{-1} \\ &= C_0^{-1} - R_t Q_t R_t^T, \end{aligned}$$

where we have abbreviated

$$R_t = C_0^{-1}AU_{t-dt}$$

and

$$Q_t = (D_{t-dt}^{-1} + U_{t-dt}^T A^T C_0^{-1}AU_{t-dt})^{-1}.$$

Now plug this into the covariance update:

$$\begin{aligned} C_t &= \left[(AC_{t-dt}A^T + \sigma^2 dtI)^{-1} + B^T W^{-1} B \right]^{-1} \\ &= [C_0^{-1} - R_t Q_t R_t^T + B^T W^{-1} B]^{-1}. \end{aligned}$$

We see that the update is of low-rank form. To apply Woodbury again, we just need to simplify the term $-R_t Q_t R_t^T + B^T W^{-1} B$. Choose an orthogonal basis

$$O_t = \text{orth}([R_t \ B])$$

and then write

$$-R_t Q_t R_t^T + B^T W^{-1} B = O_t M_t O_t^T,$$

with

$$M_t = -O_t^T R_t Q_t R_t^T O_t + O_t^T B^T W^{-1} B O_t.$$

Now, finally, apply Woodbury again:⁵

$$\begin{aligned} C_t &= [C_0^{-1} - R_t Q_t R_t^T + B^T W^{-1} B]^{-1} \\ &= [C_0^{-1} + O_t M_t O_t^T]^{-1} \\ &= C_0 - C_0 O_t (M_t^{-1} + O_t^T C_0 O_t)^{-1} O_t^T C_0. \end{aligned} \tag{5}$$

We obtain U_t and D_t by truncating the SVD of the expression on the right-hand side: in Matlab, for example, do

$$[U', D'] = \text{svd}(C_0 O_t (M_t^{-1} + O_t^T C_0 O_t)^{-1/2}, 'econ'),$$

then choose U_t as the first k columns of U' and D_t as the negative square of the first k diagonal elements D' , where k is chosen to be large enough (for accuracy) and small enough (for computational tractability). We have found that a reasonable choice of k is as the least solution of the inequality:

$$\sum_{i \leq k} D_{ii}^2 \geq c \sum_i D_{ii}^2; \tag{6}$$

i.e., choose k to capture at least a large fraction c of the variance in the right-hand term perturbing C_0 in Eq. (5). Figure 2 illustrates the accuracy of this approximation for the full C_t .

Now the update for μ_t is relatively standard:

$$\begin{aligned} \mu_t &= C_t \left[(AC_{t-dt}A^T + \sigma^2 dtI)^{-1} m_t + B^T W^{-1} (y_t - \mu_t^y) \right] \\ &= (P_t^{-1} + B^T W^{-1} B)^{-1} [P_t^{-1} m_t + B^T W^{-1} (y_t - \mu_t^y)] \\ &= (P_t - P_t B^T (W + B P_t B^T)^{-1} B P_t) \\ &\quad \times [P_t^{-1} m_t + B^T W^{-1} (y_t - \mu_t^y)] \\ &= m_t - P_t B^T (W + B P_t B^T)^{-1} B [s_t + m_t] + s_t, \end{aligned}$$

where we have made the abbreviations

$$P_t = C_0 + AU_{t-dt}D_{t-dt}U_{t-dt}^T A^T, \tag{7}$$

$$m_t = A\mu_{t-dt} + i_t dt, \tag{8}$$

and

$$s_t = P_t B^T W^{-1} (y_t - \mu_t^y).$$

⁵It is well-known that the Woodbury formula can be numerically unstable when the observation covariance W is small (i.e., the high-SNR case). Our applications here concern the low-SNR case instead, and therefore we have not had trouble with these numerical instabilities. However, it should be straightforward to derive a low-rank square-root filter (Howard and Jebara 2005; Treebushny and Madsen 2005; Chandrasekar et al. 2008) to improve the numerical stability here, if needed.

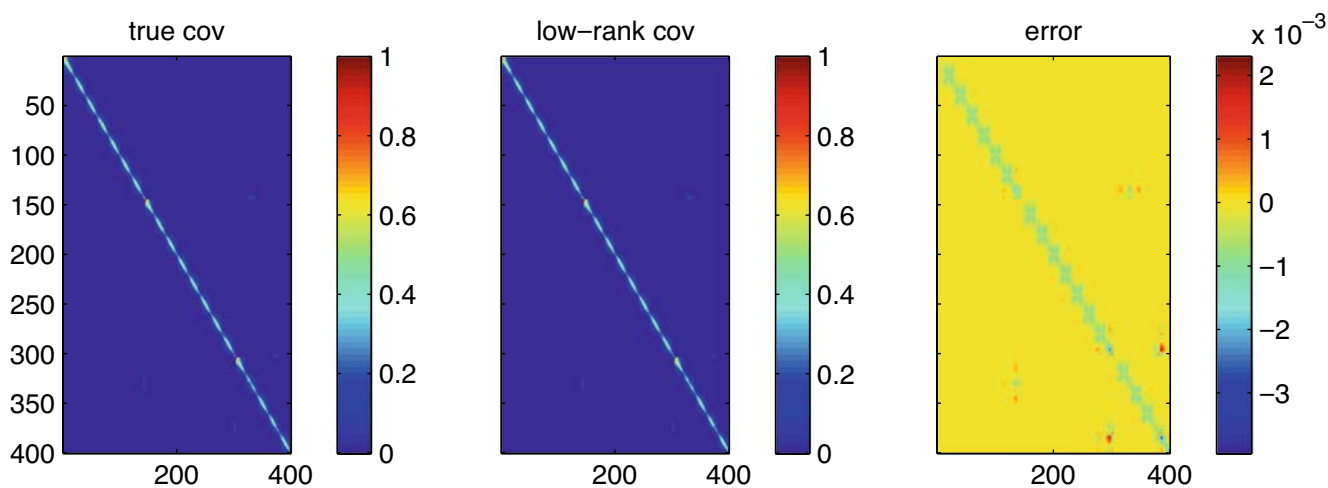


Fig. 2 C_t is very well-approximated by $C_0 + U_t D_t U_t^T$; $\text{rank}(U) = 39$ here (chosen automatically by the algorithm using $c = .999$ in Eq. (6)). This tree has two branches (near compartments 150 and 300); otherwise simulation details are as in Fig. 1. U and

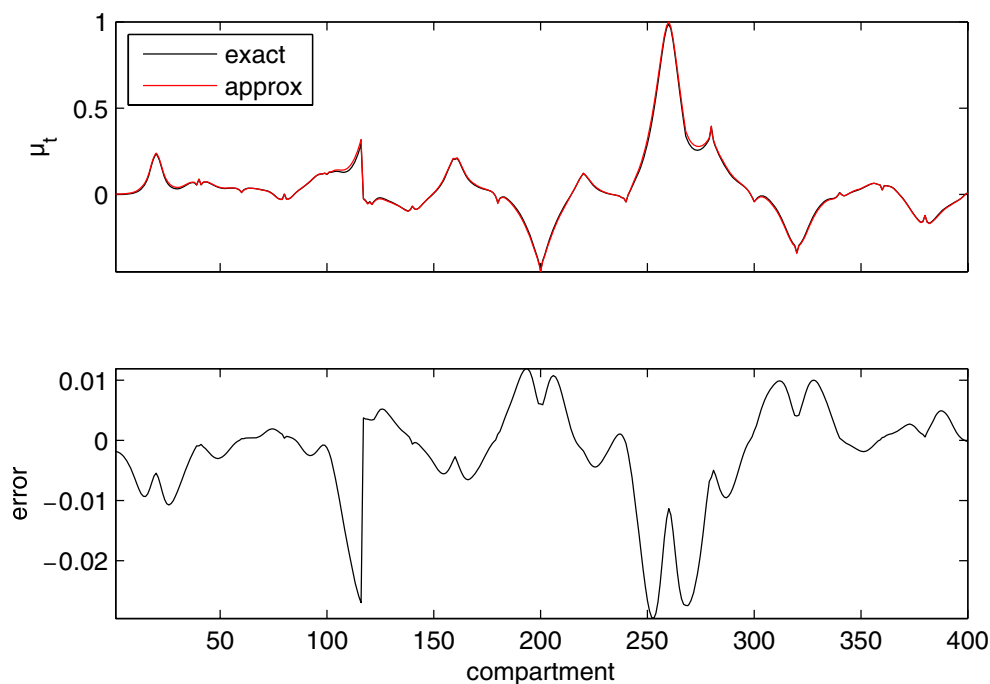
D denote the large- t limits of U_t and D_t , respectively. *Left:* true C_t . *Middle:* $C_0 + U D U^T$. *Right:* error of the approximation, $C_t - (C_0 + U D U^T)$. As in Fig. 2, each panel has been divided by $\max(C_0)$, to improve visibility

Note that we update the mean μ_t first, then truncate U_t and D_t . See Fig. 3 for an illustration.

To sum up, the key point is that all of the necessary computations described above can be done in $O(N)$ time and memory. C_0 need never be computed explicitly; instead, all we need is to multiply and divide by C_0^{-1} (by “divide,” we mean to solve equations of the form $C_0^{-1}v = r$ for the unknown vector v and known vector

r); the former can be done directly with a sparse matrix multiply, and the latter by the junction tree algorithm, or with, e.g., $C_0^{-1} \setminus r$ in Matlab (where C_0^{-1} is represented as a sparse matrix), which has efficient elimination tree code built in to its sparse matrix solver. The posterior marginal variance $[C_t]_{ii}$ can also easily be computed in $O(N)$ time, to provide errorbars around μ_t ; given the diagonal of C_0 (obtained via the junction tree

Fig. 3 The low-rank approximation for μ_t is similarly accurate. *Top:* true μ_t (black) and approximate μ_t (red) at an arbitrarily chosen timepoint t (the approximation quality remains roughly constant over the course of the trial; data not shown). *Bottom:* error of the approximation, $\mu_t - \mu_t^{\text{approx}}$. As in Fig. 2, each panel has been divided by $\max(|\mu_t|)$, to improve visibility. Note that the error is small; the details of μ_t here are driven by noise and are relatively unimportant in this example. All parameters of the simulation are as in Fig. 2



algorithm, as discussed above), computing the diagonal of C_t just requires us to sum the squared elements of $(-D_t)^{1/2}U_t$.

We should also note that the method can be sped up significantly in the special case that B is time-invariant: in this case, C_t will converge to a limit (as an approximate solution of the corresponding Riccati equation), and we can stop recomputing U_t and D_t on every time step. More generally, if B_t is time-varying in a periodic manner (e.g., we are repeatedly sequentially scanning over a region of dendrite in an imaging experiment), then C_t will also be periodic, and we can store a period's worth of U_t and D_t in memory, instead of continuing to recompute these on each timestep.

2.1 Incorporating implicit methods for solving the cable equation

It is well-known that the forward Euler method described above for solving the cable equation is unstable for large values of adt (where a denotes the inter-compartmental conductance) (Hines 1984; Press et al. 1992). The following “backwards Euler” (implicit) discretization is unconditionally stable and can therefore be made much faster, by choosing a larger timestep dt :⁶

$$V_{t+dt}(x) = V_t(x) + dt \left(-g_x V_{t+dt}(x) + \sum_{w \in N(x)} a_{xw} [V_{t+dt}(w) - V_{t+dt}(x)] \right).$$

We can write this in matrix-vector form as

$$V_{t+dt} = V_t + R V_{t+dt}$$

for a suitable Hines matrix R , or in other words

$$V_{t+dt} = (I - R)^{-1} V_t.$$

It is conceptually straightforward to replace A with $(I - R)^{-1}$ in our above derivations. Of course, $(I - R)^{-1}$ is not a sparse matrix, so we do not want to make this substitution explicitly; instead, we will use our efficient matrix-divide methods to solve for any term of the form $(I - R)^{-1}v$. In particular, we need to multiply and divide by C_0 . We have

$$C_0^{-1} = \frac{1}{\sigma^2 dt} [I - A^2] = \frac{1}{\sigma^2 dt} [I - (I - R)^{-2}];$$

⁶We have suppressed the input current $i_t(x)$ and noise $\epsilon_t(x)$ in the following, for notational simplicity; inclusion of the input current term does not significantly complicate the derivations in this section, since the steady-state covariance C_0 does not depend on $i_t(x)$.

$(I - R)^2$ has tree-bandwidth two (i.e., matrix elements corresponding to pairs of compartments which are further than two nodes apart on the tree are zero), so we can efficiently divide by this matrix. Similarly,

$$\begin{aligned} C_0 &= \sigma^2 dt [I - A^2]^{-1} \\ &= \sigma^2 dt [I - (I - R)^{-2}]^{-1} \\ &= \sigma^2 dt [(I - R)^2 - I]^{-1} (I - R)^2; \end{aligned}$$

multiplying by $(I - R)^2$ is clearly fast, and since $[(I - R)^2 - I]$ has tree-bandwidth two, we can efficiently divide by this matrix as well.

Second-order semi-implicit Crank-Nicholson methods (Hines 1984; Press et al. 1992) may be incorporated in a similar fashion, though we have found the fully implicit method described here to be adequate for our needs, and have not pursued the Crank-Nicholson implementation in depth.

At this point it is worth pausing to summarize our approach. Given the reconstructed neuron file (containing the neighborhood structure $N(x)$) and the observed matrix of voltage-sensitive measurements Y , the first step is to define the parameters in the dynamics and observations Eqs. (1) and (2). Specifically, we must form the dynamics matrix A , the input current vector i_t , the observations matrix B_t , and the noise parameters σ^2 , W_t , and μ_t^y . The observation parameters B_t , W_t , and μ_t^y are characteristics of the imaging apparatus, and good initial estimates of these parameters should therefore be available directly: as discussed above, the i -th row of B_t specifies how the observed voltage at each compartment is weighted and summed to form the observations at the i -th pixel at time t , while W_t , and μ_t^y specify the noise covariance and bias of these observations, respectively. The dynamics parameters A , i_t , and σ can also be initialized in a straightforward manner: in the simplest case, we could let the intercompartmental couplings a_{xw} and membrane conductance g_x be proportional to the compartment diameter, reducing the difficult task of choosing A to the much simpler job of selecting two proportionality coefficients. As discussed in Huys and Paninski (2009), these parameters (along with the dynamics noise variance σ^2) may be estimated using an iterative expectation-maximization approach, and initialized using prior knowledge from previous estimates in similar neurons. Once all of these system parameters are chosen, we need only select the fraction of variance parameter c (from Eq. (6)): larger values of c lead to greater accuracy at the cost of speed. Then we simply iterate the update equations above to obtain the desired filter outputs $E(V_t|Y_{1:t})$

and $Cov(V_t|Y_{1:t})$. Sample code implementing the simulations described in the next section is available at <http://www.stat.columbia.edu/~liam/research/abstracts/dendrite-kalman-abs.html>.

3 Applications to spatially-sampled dendritic trees

We illustrate the fast implicit Kalman filtering methods described in the last section with two simulations here. The results of the simulations are best viewed in movie form (low-rank-speckle.mp4 and low-rank-horiz.mp4); we show snapshots and two-dimensional representations of these movies in Figs. 4, 5, 6, 7, and 8. In the first three figures, we demonstrate the output of the fast Kalman filter given sparsely-sampled spatial observations of a complex, real dendritic tree (reconstructed from a rabbit starburst amacrine cell), and in the last two figures we apply the fast Kalman filter to highly spatially-coarsened observations: we observe only the vertically-summed, noise-corrupted voltage on the tree,

instead of retaining full spatial information about the observations Y . (This latter simulation is an effort to roughly emulate the results of fast planar imaging methods, as discussed for example in Holekamp et al. (2008).)

In each case, the filter is able to recover the spatiotemporal voltage fairly well (and is even able to restore some of the spatial information that is apparently lost by vertically summing the observations in Figs. 7–8), though the filtered voltage $E(V_t|Y_{1:t})$ consistently underestimates and smooths the true voltage V_t , as we would expect of any optimal linear filter. We emphasize that our goal here is to give a qualitative picture of the filter's behavior, to illustrate that the approximate methods developed above do not lead to any visible artifacts in the filter output; for more quantitative analyses of the estimation performance of the Kalman and other Bayesian filters applied in this setting (albeit in the context of much lower-dimensional neural models, where more direct computational approaches are tractable), see Huys and Paninski (2009).

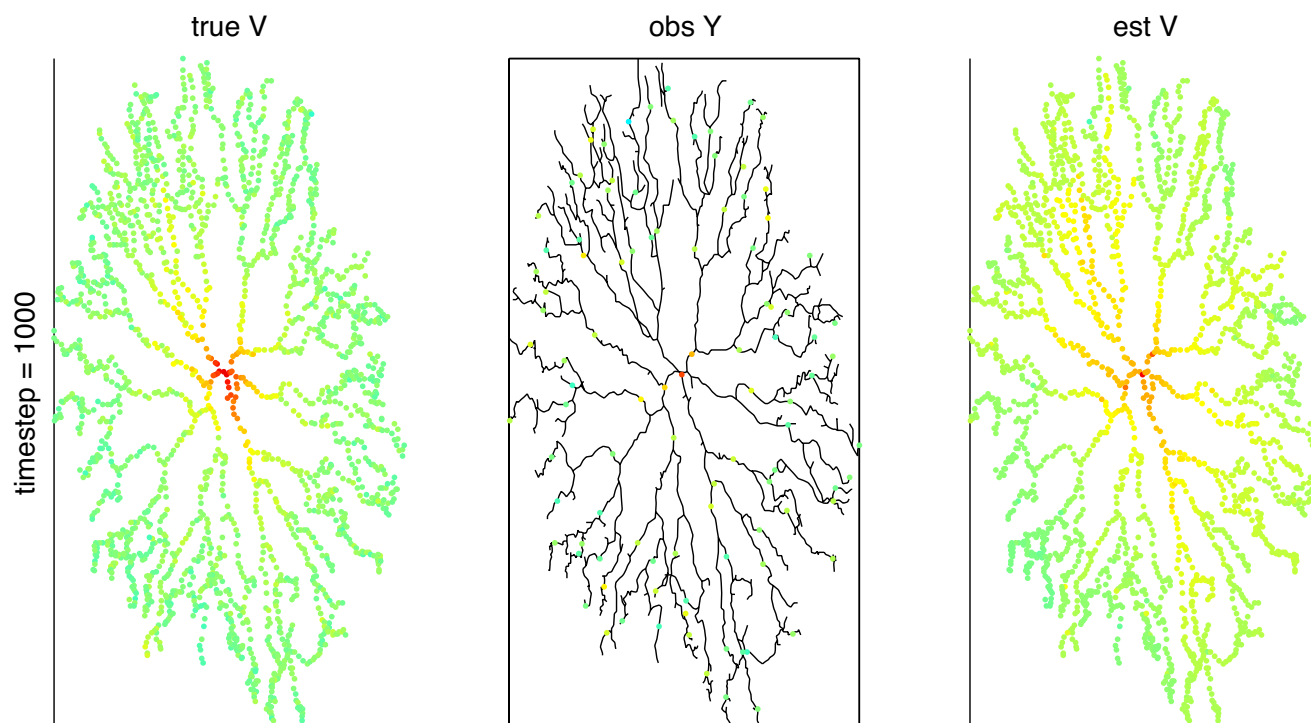


Fig. 4 Snapshot of a simulation in which we made noisy observations of just 100 out of > 2000 compartments. The full simulation may be viewed in movie form in low-rank-speckle.mp4. The neuronal geometry is taken from the “THINSTAR” file (rabbit starburst amacrine cell) available at <http://neuromorpho.org>; see Bloomfield and Miller (1986) for details. *Left panel*: true voltage $V_t(x)$ at timestep $t = 1000$; color indicates voltage (red corresponds to high voltage and blue corresponds to low). Each dot represents an individual compartment. The somatic compartment

(center, red) was driven with a sinusoidal input (see Fig. 5 below for details). *Middle panel*: observed data y_t . 100 compartments were randomly chosen and then observed (with noise) at each time step. *Underlying black trace* indicates structure of dendritic tree. Each colored dot indicates an observed compartment. *Right panel*: filter estimate $E(V_t(x)|Y_{1:t})$. The first two panels are plotted on the same color scale; the third is plotted on its own color scale for improved visibility. See Fig. 5 for the full $V_t(x)$, y_t , and $E(V_t(x)|Y_{1:t})$ for all times t

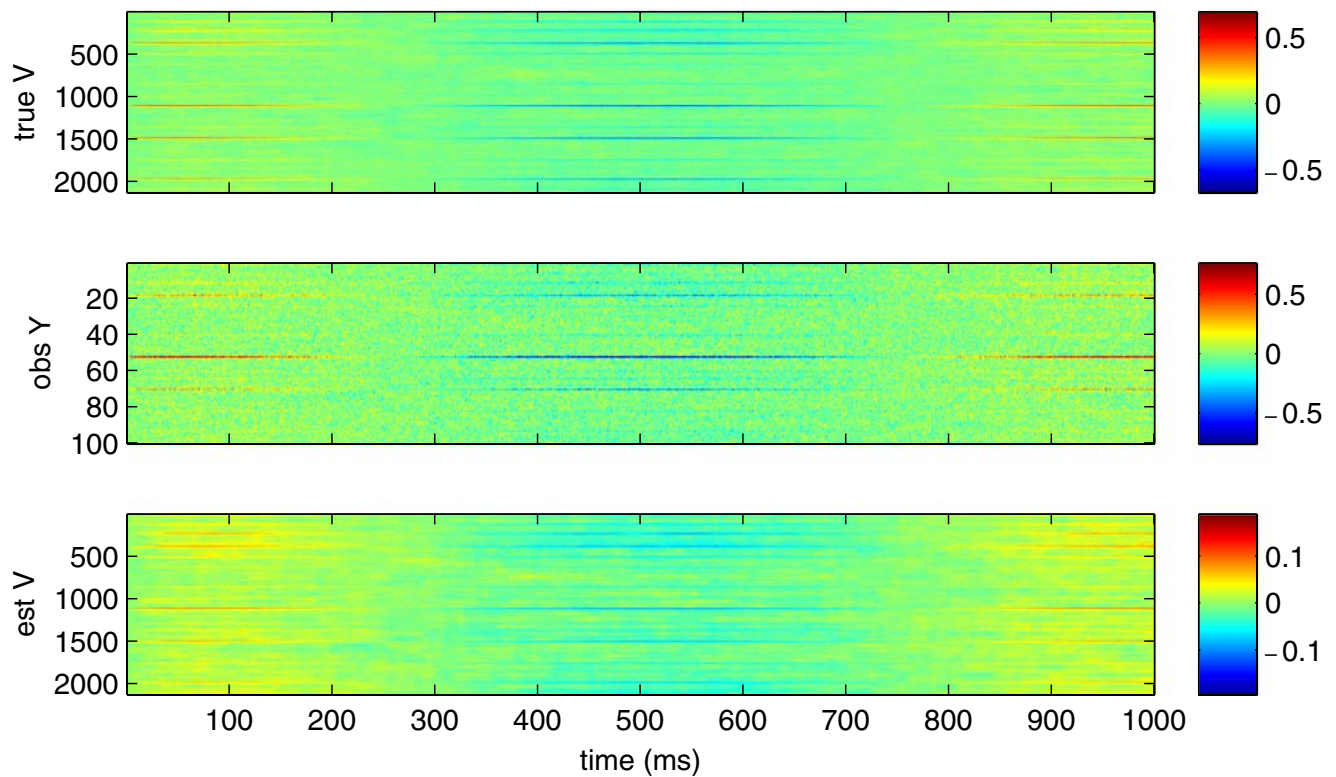


Fig. 5 Full spatiotemporal voltage traces corresponding to the snapshots shown in Fig. 4; data matrices may be obtained by running the code at <http://www.stat.columbia.edu/~liam/research/abstracts/dendrite-kalman-abs.html>. *Top*: true spatiotemporal voltage $V_t(x)$; *middle*: observed data y_t ; *bottom*: filtered voltage $E(V_t(x)|Y_{1:t})$. All plots share the same units (voltage scaled by a factor which in this case may be ignored), but are plotted with different *color scaling*, for improved visibility. In this simulation, a 1 Hz sinusoidal current was injected into compartment 1 (visible

as the soma, colored *red* in the *left panel* in Fig. 4); however, this current was not considered known (i.e., i_t in Eq. (1) was set to zero), to make the filtering problem a bit more challenging here. The average SNR per compartment (computed by taking the average signal variance $var(V_t(x))$ over all the compartments and dividing by the observation noise variance, normalized by the fraction of observed compartments, n/N) was about 4%. Note that the filtered voltage $E(V_t|Y_{1:t})$ underestimates and smooths the true voltage V , both spatially and temporally

4 Extending the basic method

In this section we discuss several important extensions of the basic methods introduced in Section 2.

4.1 Temporally-filtered observations and dynamics noise

First we would like to relax the assumption that the observations y_t are instantaneously related to the voltage V_t . For example, it is well-known that whole-cell recordings introduce filtering effects, due to the RC dynamics of the electrode-cell system (Brette et al. 2007); similarly, some voltage sensors (particularly genetically-encoded voltage sensors) respond non-instantaneously to voltage steps and thus effectively lowpass filter the voltage (Knopfel et al. 2006).

We therefore modify our basic dynamics and observation equations to incorporate this temporal filtering

of the observations. For clarity, we will treat just the simplest case, in which the observations are filtered in a first-order (low-pass) manner, though generalizations are straightforward. We write

$$q_{t+dt} = Aq_t + \sqrt{dt}\epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, I)$$

$$z_{t+dt}^i = z_t^i + \left(aq_t^i - \frac{z_t^i}{\tau} \right) dt + \sqrt{dt}\xi_t^i, \quad \xi_t^i \sim \mathcal{N}(0, \sigma^2)$$

$$y_t = Bz_t + \eta_t, \quad \eta_t \sim \mathcal{N}(0, W).$$

Here we have introduced a scaled voltage variable q_t to simplify the notation below, and we make linear observations of the low-pass filtered variable z_t ; τ denotes the time constant of this filter, and a denotes the filter gain. (A minor technical point: since we do not observe q_t directly, there is a free scale parameter which must be fixed here; thus we have set the q -dynamics noise variance to 1, and set the z -dynamics noise variance to σ^2 . We have also set the observation

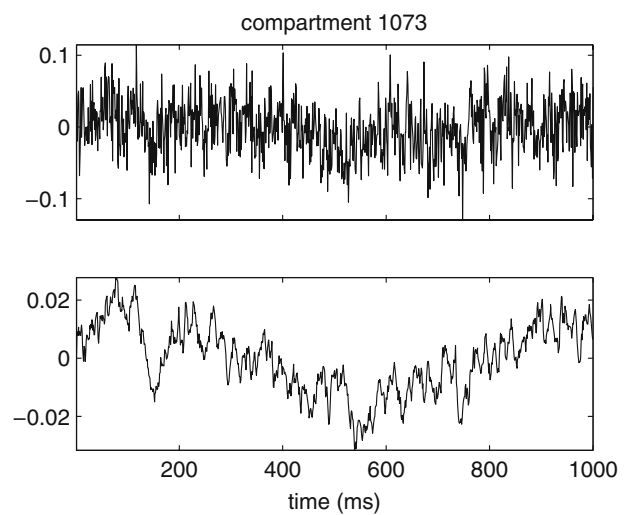
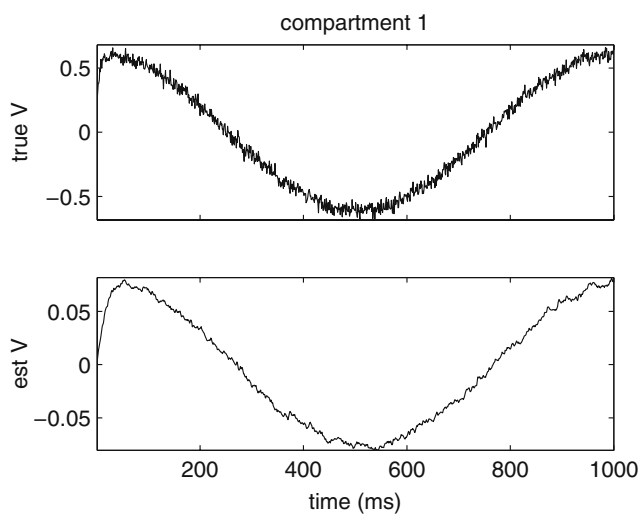


Fig. 6 Example one-dimensional voltage traces selected from the data matrices shown in Fig. 5. The true and inferred somatic voltages are shown on the *left* (compartment 1); voltages from a randomly-chosen compartment are shown on the *right*. Neither of these compartments were among the subset of 100 compartments which were observed directly in the middle panels of

Figs. 4–5. The effect of the sinusoidal somatic input current is most strongly visible on the *left*, and is more attenuated in the distal compartment shown on the *right*. Again, note that the filtered voltage significantly underestimates and smooths the true voltage V , due to the low effective SNR here

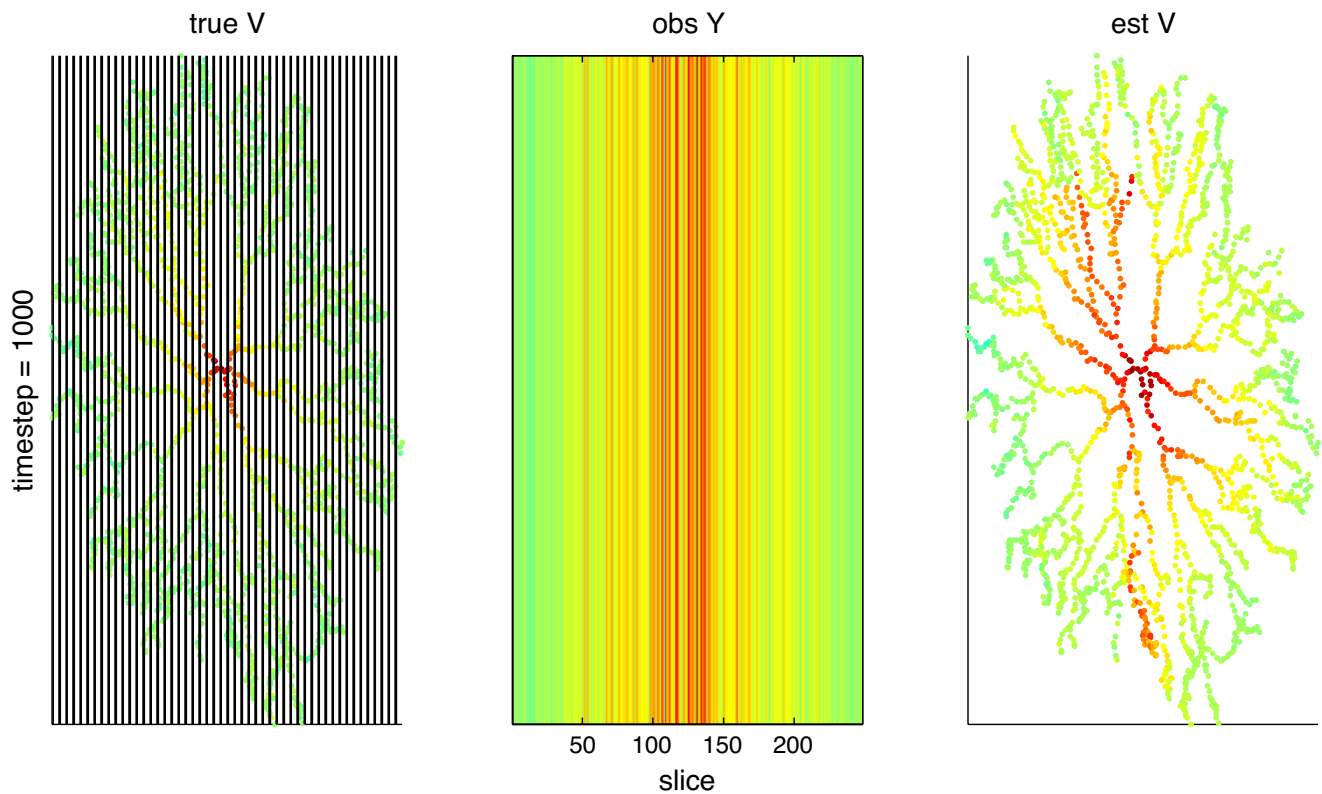


Fig. 7 Snapshot of a simulation similar to that shown in Figs. 4–5, but with vertically-scanned observations (i.e., the dynamics equation is identical here, but the observation equation is different). Layout and conventions are as in Fig. 4. In this case, the

observations Y were formed by summing the observed voltages over the 250 nonoverlapping “slices” shown in the *left panel* (only every fifth slice is shown, for visibility) and adding noise. The full simulation may be viewed in movie form in [low-rank-horiz.mp4](#)

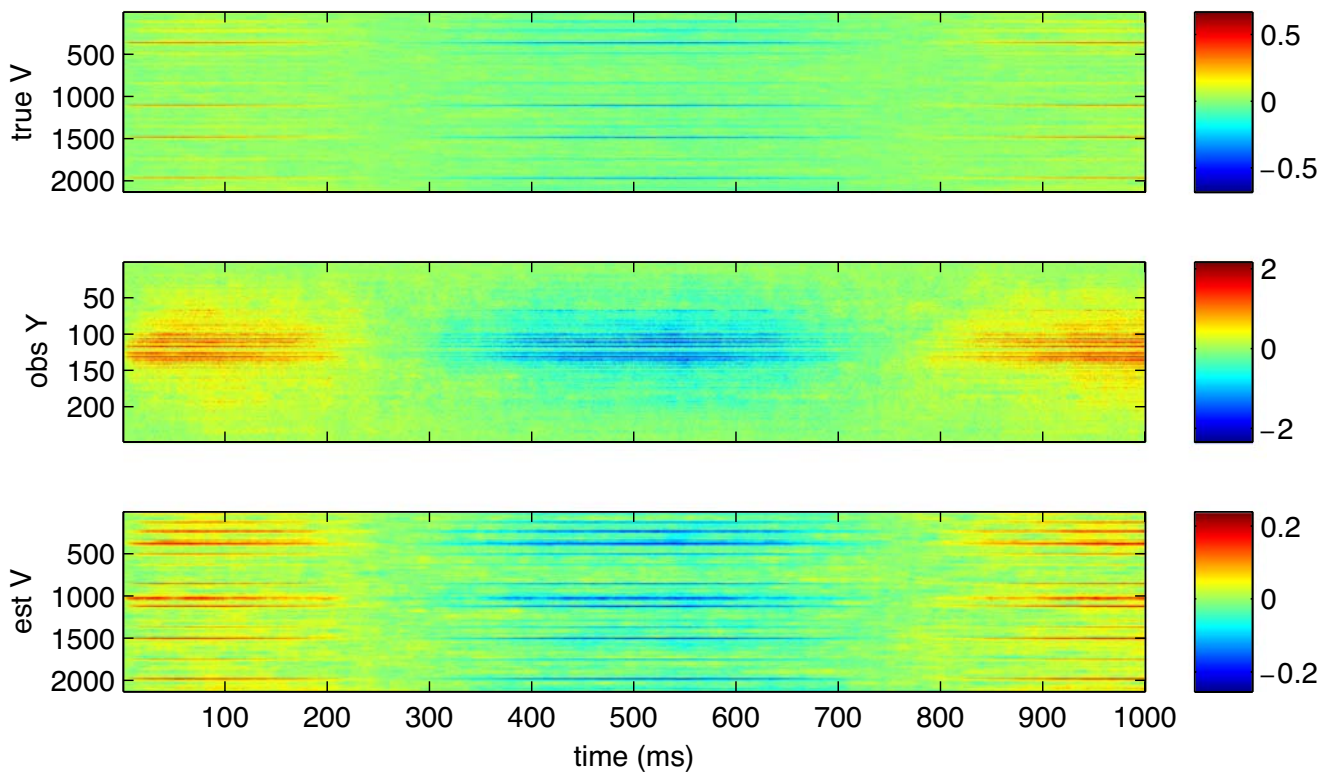


Fig. 8 Full spatiotemporal voltage traces corresponding to the snapshots shown in Fig. 7. As in Figs. 4–5, sinusoidal current was injected into the soma but not included in the Kalman filter ($i_t = 0$). Again, the filtered voltage $E(V_t|Y_{1:t})$ significantly

smooths and shrinks the true voltage V . In addition, the vertical sampling here leads to significant loss of spatial detail relative to the simulation shown in Figs. 4–5

mean μ_t^y to zero, and suppressed the input current $i_t(x)$ for notational simplicity; inclusion of these terms does not significantly complicate the following derivations, since again the steady-state covariance C_0 does not depend on μ_t^y or $i_t(x)$.

Now the key to the basic methods described in Section 2 is that we could compute the steady-state covariance C_0 of the state vector q_t quite explicitly. In the case of interest here, the state vector is (q_t, z_t) ; unfortunately, computing the steady-state covariance (which we will continue to denote as C_0) is a bit less trivial now, since the dynamics matrix is asymmetric, and therefore the infinite sum Eq. (4) is more difficult to compute. We write

$$C_0 = \begin{pmatrix} c_0 & J \\ J^T & F \end{pmatrix},$$

where we have abbreviated the upper-left block as

$$c_0 = (I + A^2)^{-1} dt$$

(i.e., c_0 denotes the familiar steady-state covariance of q_t).

To compute the matrices J and F , we rewrite the z_t dynamics slightly,

$$z_{t+dt} = b z_t + dq_t + \sqrt{dt}\xi_t, \quad \xi_t \sim \mathcal{N}(0, \sigma^2 I),$$

where we have abbreviated $b = 1 - dt/\tau$ and $d = adt$. Now we can write a simple recursion for the cross-covariance:

$$\begin{aligned} E(q_{t+dt}z_{t+dt}^T) &= E\left(\left[Aq_t + \sqrt{dt}\epsilon_t\right]\left[bz_t + dq_t + \sqrt{dt}\xi_t\right]^T\right) \\ &= dAE(q_tq_t^T) + bAE(q_tz_t^T). \end{aligned}$$

Here $E(q_tq_t^T)$ tends to c_0 , as before, and $E(q_{t+dt}z_{t+dt}^T)$ should be equal to $E(q_tz_t^T)$ in the limit $t \rightarrow \infty$. This leads to the equation

$$J = d(I - bA)^{-1}Ac_0.$$

Note that J is symmetric, since it is a product of symmetric commuting terms.

We may derive a similar recursion for the covariance of z_t :

$$cov(z_{t+dt}) = b^2cov(z_t) + d^2cov(q_t) + 2bdE(q_tz_t^T) + \sigma^2 Idt.$$

Taking limits as above, we compute

$$F = (1 - b^2)^{-1} (d^2 c_0 + 2bdJ + \sigma^2 Idt).$$

Now we need to efficiently multiply and divide by C_0 . Multiplying by C_0 can clearly be done in $O(N)$ time. To divide, we write the inverse as

$$C_0^{-1} = \begin{pmatrix} c_0^{-1} + c_0^{-1} J G^{-1} J c_0^{-1} & -G^{-1} J c_0^{-1} \\ -G^{-1} J c_0^{-1} & G^{-1} \end{pmatrix},$$

where we have abbreviated the Schur complement

$$G = F - J c_0^{-1} J.$$

Multiplying by J and c_0^{-1} can clearly be done in $O(N)$ time. After some cancellations, G may be written as

$$G = \left[(1 - b^2)^{-1} (d^2 (I - b^2 A^2) + \sigma^2 dt (I - b A)^2 c_0^{-1}) - d^2 A^2 \right] c_0 (I - b A)^{-2}.$$

The term in brackets has a tree-bandwidth of 4, and we can therefore divide by G in $O(N)$ time, as desired. See Fig. 9 for an illustration.

A very similar approach may be used to handle the case of temporally-correlated input noise, e.g.,

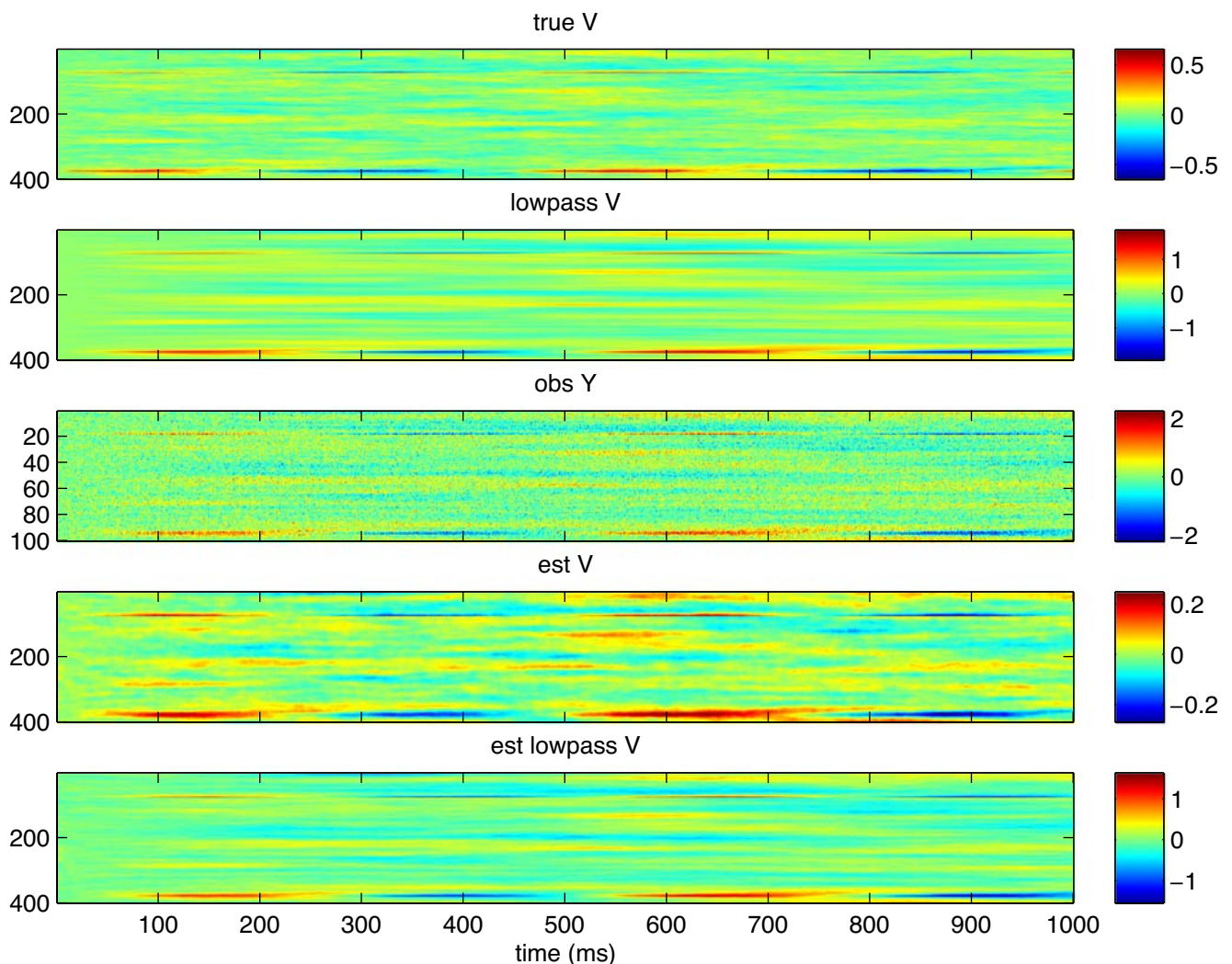


Fig. 9 An example of the Kalman filter applied to lowpass-filtered voltage observations. *Top panel:* true spatiotemporal voltage. Sinusoidal current (2 Hz) was injected into a branch point of the tree (compartment 380); the tree was a simple one-branched structure with 400 compartments in this case. *Second panel:* lowpass-filtered ($\tau = 100$ ms) voltage. *Third panel:* observed Y ; the lowpass-filtered voltage was spatially subsampled

4x and corrupted with noise. Fourth and fifth panels: output of Kalman filter: estimated voltage V and lowpass-filtered voltage, respectively. Note that the Kalman filter is causal, so the estimated voltage V lags the true V ; to obtain a non-lagged estimate, we must run the full forward-backward Kalman smoother (data not shown)

synaptically-filtered noise arising from spontaneous release of neurotransmitter, followed by postsynaptic conductance transients which do not decay instantaneously. As above, we treat the case of a first-order lowpass filter, and begin with the recursion

$$q_{t+dt} = Aq_t + dw_t + \sqrt{dt}\epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, I)$$

$$w_{t+dt} = bw_t + \sqrt{dt}\xi_t, \quad \xi_t \sim \mathcal{N}(0, \sigma^2 I),$$

where w_t represents the filtered noise source. From here we may derive a simple recursion for the cross-covariance:

$$\begin{aligned} E(q_{t+dt}w_{t+dt}^T) &= E\left(\left[Aq_t + dw_t + \sqrt{dt}\epsilon_t\right]\left[bw_t + \sqrt{dt}\xi_t\right]^T\right) \\ &= bAE(q_t w_t^T) + b dE(w_t w_t^T). \end{aligned}$$

Computing the steady-state solution of this recursion, we find that the cross-covariance $E(q_t w_t^T)$ tends to

$$J = b d \sigma^2 (1 - b^2)^{-1} (I - bA)^{-1}.$$

We also have that

$$\begin{aligned} E(q_{t+dt}q_{t+dt}^T) &= AE(q_t q_t^T) A^T + d^2 E(w_t w_t^T) \\ &\quad + 2dAE(q_t w_t^T) + Idt, \end{aligned}$$

from which we conclude that the covariance $E(q_t q_t^T)$ tends to

$$\begin{aligned} F &= (I - A^2)^{-1} [(d^2 \sigma^2 (1 - b^2)^{-1} + dt)I \\ &\quad + 2bd^2 \sigma^2 (1 - b^2)^{-1} A(I - bA)^{-1}]. \end{aligned}$$

The full steady-state covariance

$$C_0 = \begin{pmatrix} (1 - b^2)^{-1} \sigma^2 I & J \\ J & F \end{pmatrix}$$

may now be multiplied and divided in $O(N)$ time using sparse Schur methods nearly identical to those described above.

4.2 Inhomogeneous cable noise and quasi-active membranes

In the case that the dynamics noise covariance $cov(\epsilon)$ is diagonal but not proportional to the identity, the infinite sum Eq. (4) is even more difficult to compute (since $cov(\epsilon)$ and A will not commute in general). Standard direct methods for solving the Lyapunov equation (e.g., the Bartels-Stewart algorithm (Antoulas 2005)) require an orthogonalization step that takes $O(N^3)$ time in general. There is a large applied mathematics literature on the approximate solution of Lyapunov equations with sparse dynamics (see e.g. Sabino (2007) for a nice review), but the focus of this literature is on the case that the noise covariance matrix is of low rank.

Since we expect the voltage at every compartment to be perturbed by at least some noise (implying that the rank of the noise covariance is at least as large as N), these low-rank methods are less attractive here. We use a direct iterative method for computing C_0 instead.

Abbreviate $D = cov(\epsilon)$. Now we recurse

$$\begin{aligned} S_{i+1} &= (AS_i^{-1}A + D)^{-1} \\ &= D^{-1} - D^{-1}A(S_i + AD^{-1}A)^{-1}AD^{-1}; \end{aligned} \quad (9)$$

the solution of this recursion is exactly the steady-state inverse covariance C_0^{-1} . The key now is that the iterates S_i remain “almost” tree-banded if the elements of D vary smoothly along the tree; i.e., we can safely set all but $O(N)$ of the elements of S_i to zero without incurring much numerical error.⁷ In particular, we find that the limit $S_\infty = C_0^{-1}$ remains effectively tree-banded in this case.

Now, if S_i is tree-banded, then so is $(S_i + AD^{-1}A)$, as is the Cholesky decomposition L_i defined by

$$L_i L_i^T = (S_i + AD^{-1}A).$$

Similarly, since S_{i+1} and D^{-1} are both tree-banded, then so is $L_i^{-1}AD^{-1}$, and therefore we can compute this term—and from here complete the recursion for S_{i+1} —in $O(N)$ time. Once the limit of this recursion (i.e., our approximation to C_0^{-1}) is obtained, we may apply the $O(N)$ methods described in Section 2 with no further modifications. We have found that a very effective initialization for this procedure is to set

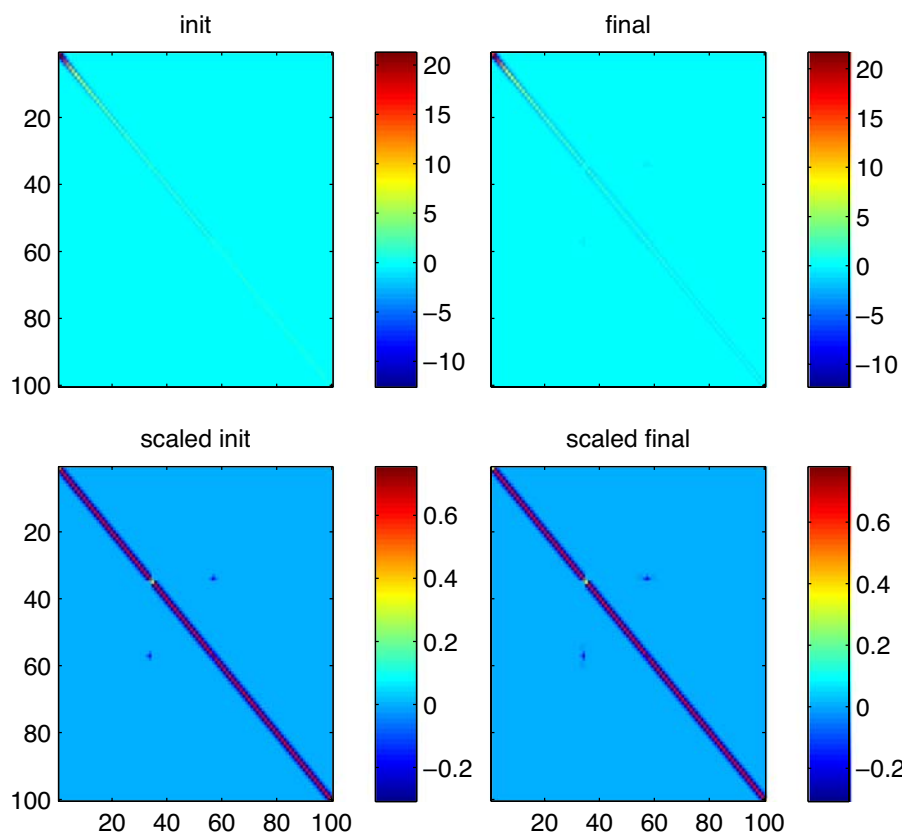
$$S_0 = D^{-1/2}(I - A^2)D^{-1/2}; \quad (10)$$

this may be computed in $O(N)$ time and gives the exact solution C_0^{-1} in the case that D is proportional to the identity. See Fig. 10 for an illustration.

A very similar problem arises in the case of “quasi-active” membranes (Koch 1984; Manwani and Koch 1999; Coombes et al. 2007; Kellems et al. 2009), in which the dynamics of each compartment are described by a k -dimensional linear system (instead of the one-dimensional stable dynamics for each compartment we have used in the rest of this paper); this linear system is derived from a first-order expansion about a fixed point of the original nonlinear model neuronal dynamics. In this setting the dynamics matrix A may be written most conveniently as a block-tree-tridiagonal matrix, with block size equal to k ; see, e.g., Kellems et al. (2009) for

⁷We do not have a good theoretical explanation for this fact; this is a purely empirical observation. Indeed, we have found that if the noise scale D varies discontinuously along the tree, then S_i tends to be much less sparse, sharply reducing the efficacy of this procedure.

Fig. 10 Computing C_0^{-1} via the sparse recursion Eq. (9). In this case, the tree was a simple single-branched structure (with the branch at compartment 35), and the noise variance D varied linearly over two orders of magnitude as a function of the compartment number. *Left:* initial approximate inverse covariance S_0 (top; Eq. (10)) and rescaled matrix $D^{1/2}S_0D^{1/2}$ (bottom), included for clarity. *Right:* limit matrix $S_\infty = C_0^{-1}$ (top) and rescaled matrix $D^{1/2}S_\infty D^{1/2}$ (bottom). Note that S_0 provides a very good initialization, and that both S_0 and C_0^{-1} are very sparse



further details. The problem is that A here will be asymmetric in general, and therefore, once again, we can not solve the Lyapunov equation directly by computing the infinite matrix sum Eq. (4) explicitly. However, we have found that this quasi-active case may be handled in a very similar manner; we simply recurse Eq. (9), with S_0 set to $D^{-1/2}(I - AA^T)D^{-1/2}$ (Eq. (10) needs to be modified slightly since A is asymmetric here). As in the asymmetric-noise case discussed above, we find empirically that S_i remains tree-banded, making the recursion Eq. (9), and therefore the computation of C_0^{-1} , quite tractable.

4.3 Nonlinear voltage or calcium observations

The last extension we consider involves the case of nonlinear observations y_t . For example, some voltage sensors saturate somewhat over their dynamic range, rendering the linear model $y_t = B_t V_t + \eta_t$ inadequate. It is also worth considering observations of fluorescent calcium sensors: as discussed above, voltage-sensing techniques are currently hampered by low SNR (although optimal filtering methods can at least partially alleviate this problem). On the other hand, spatiotemporal calcium-sensing fluorescence signals may be cur-

rently recorded at relatively high SNR (Gobel and Helmchen 2007; Larkum et al. 2008), but calcium signals, by their nature, provide only highly-thresholded information about the underlying voltage, since calcium channels are inactive at hyperpolarized voltages. Can we exploit these high-SNR superthreshold calcium observations to reconstruct (at least partially) the sub-threshold voltage signal? More generally, we would like to optimally combine calcium and voltage measurements, where voltage measurements may be available via imaging techniques, or whole-cell patch recordings at the soma or apical dendrite, or even through dense multielectrode recordings (Petrusca et al. 2007).

To incorporate calcium observations into our Kalman filtering framework, we must first write down a dynamical model for the evolution of the intracellular calcium concentration. For simplicity, we consider a first-order model for these calcium dynamics:

$$C_{t+dt}(x) = C_t(x) + \left(-\frac{C_t(x)}{\tau_C} + k \left[C_t(x+dx) - 2C_t(x) + C_t(x-dx) \right] + f[V_t(x)] \right) dt + \sigma_C \sqrt{dt} \epsilon_t(x),$$

where $C_t(x)$ denotes the calcium concentration in compartment x at time t , k is a diffusion constant, $k[C_t(x + dx) - 2C_t(x) + C_t(x - dx)]$ represents diffusion within the dendrite (this assumes that x corresponds to an interior compartment of a linear segment of the dendrite, and may be easily modified in the case of a boundary or branching compartment), and $f(V)$ is a nonlinear function corresponding to voltage-dependent calcium influx. (Note that we have approximated the calcium channel dynamics as instantaneous here, though this assumption can be relaxed.)

Now the important point is that the intracellular and transmembrane calcium diffusion terms k and $1/\tau_C$ are relatively small; in the subthreshold regime the calcium concentration changes much more slowly than the voltage. We can take advantage of this fact: if we sample sufficiently rapidly along the dendritic tree, then we may obtain $C_t(x)$ (up to some observation noise) and then numerically subtract the estimated $dC_t(x)/dt$ from the linear terms on the right hand side of our dynamics equation to obtain, finally, our observation y_t , which will correspond to a nonlinear measurement $f[V_t(x)]$ of the voltage. Of course, this will be a noisy measurement; the variance of this noise can be computed given the variance of the dynamics noise ϵ and the calcium-sensitive observation fluorescence noise. More precisely, if $y_t = BC_t + \eta_t$, and k and $1/\tau_C$ are both small, then we may approximate

$$y_{t+dt} - y_t \approx B \left(f(V_t)dt + \sigma_C \sqrt{dt} \epsilon_t \right) + \eta_{t+dt} - \eta_t.$$

If η_t and ϵ_t are approximately Gaussian, then if we define the new observation $y'_t = y_t - y_{t-dt}$, we can write

$$y'_t \approx h(V_t) + \epsilon'_t,$$

for a suitable function $h(\cdot)$ and Gaussian noise ϵ'_t with a suitable covariance matrix. (We note in passing that a similar approach may be suitable for sodium-sensitive imaging data (Kole et al. 2008).)

Thus, more generally, we would like to incorporate observations y_t obeying some arbitrary conditional density $p(y_t|V_t)$ into our filter equations. This is of course difficult in general, since if $p(y_t|V_t)$ is chosen maliciously it is clear that our posterior distribution $p(V_t|Y_{1:t})$ may be highly non-Gaussian, and our basic Kalman recursion will break down. However, if $\log p(y_t|V_t)$ is a smooth function of V_t , it is known that a Gaussian approximation to $p(V_t|Y_{1:t})$ will often be fairly accurate (Fahrmeir and Tutz 1994; Brown et al. 1998; Paninski et al. 2009), in which case our Kalman recursion may be adapted in a fairly straightforward manner.

For simplicity, we will focus on the case that the observations y_{it} are independent samples from $p(y_{it}|B_i V_t)$, where B_i denotes the i -th row of the observation matrix B . We approximate the posterior mean μ_t with the maximum a posteriori (MAP) estimate,

$$\begin{aligned} \mu_t &\approx \arg \max_{V_t} \left[\log p(V_t|Y_{0:t-dt}) + \log p(y_t|V_t) \right] \\ &\approx \arg \max_{V_t} \left[-\frac{1}{2}(V_t - m_t)^T P_t^{-1}(V_t - m_t) \right. \\ &\quad \left. + \sum_i \log p(y_{it}|B_i V_t) \right]. \end{aligned}$$

(Recall that the one-step covariance matrix P_t and mean m_t were defined in Eqs. (7–8).) This MAP update is exact in the linear-Gaussian case (and corresponds exactly to the Kalman filter), but is an approximation more generally.

To compute this MAP estimate, we use Newton’s optimization method. We need the gradient of the log-posterior with respect to V_t ,

$$\nabla = -P_t^{-1}(V_t - m_t) + B^T f_1(V_t),$$

and the Hessian with respect to V_t ,

$$H = -P_t^{-1} + B^T \text{diag}[f_2(V_t)]B.$$

Here $f_1(V_t)$ and $f_2(V_t)$ are the vectors formed by taking the first and second derivatives, respectively, of $\log p(y_{it}|u)$ at $u = B_i V_t$, with respect to u . Now we may form the Newton step:

$$\begin{aligned} V_{new} &= V_{old} - H^{-1} \nabla \\ &= V_{old} - \left(-P_t^{-1} + B^T \text{diag}[f_2(V_{old})]B \right)^{-1} \\ &\quad \times \left(-P_t^{-1}(V_{old} - m_t) + B^T f_1(V_{old}) \right) \\ &= V_{old} - \left(P_t - P_t B^T \left(-\text{diag}[f_2(V_{old})]^{-1} \right) \right. \\ &\quad \left. + B P_t B^T \right)^{-1} B P_t \left[P_t^{-1}(V_{old} - m_t) - B^T f_1(V_{old}) \right] \\ &= m_t + P_t B^T \left(-\text{diag}[f_2(V_{old})]^{-1} + B P_t B^T \right)^{-1} \\ &\quad \times B \left[V_{old} - m_t - P_t B^T f_1(V_{old}) \right] + P_t B^T f_1(V_{old}) \end{aligned}$$

We iterate, using a backstepping linesearch to guarantee that the log-posterior increases on each iteration, until convergence (i.e., when $V_{new} \approx V_{old}$, set $\mu_t = V_{old}$). Then, finally, we update the covariance C_t by replacing W^{-1} with $-\text{diag}[f_2(V_t)]$ in the derivation in Section 2. Since multiplication by P_t requires just $O(N)$ time (and we need to compute $P_t B^T$ just once per timestep), all of these computations remain tractable.

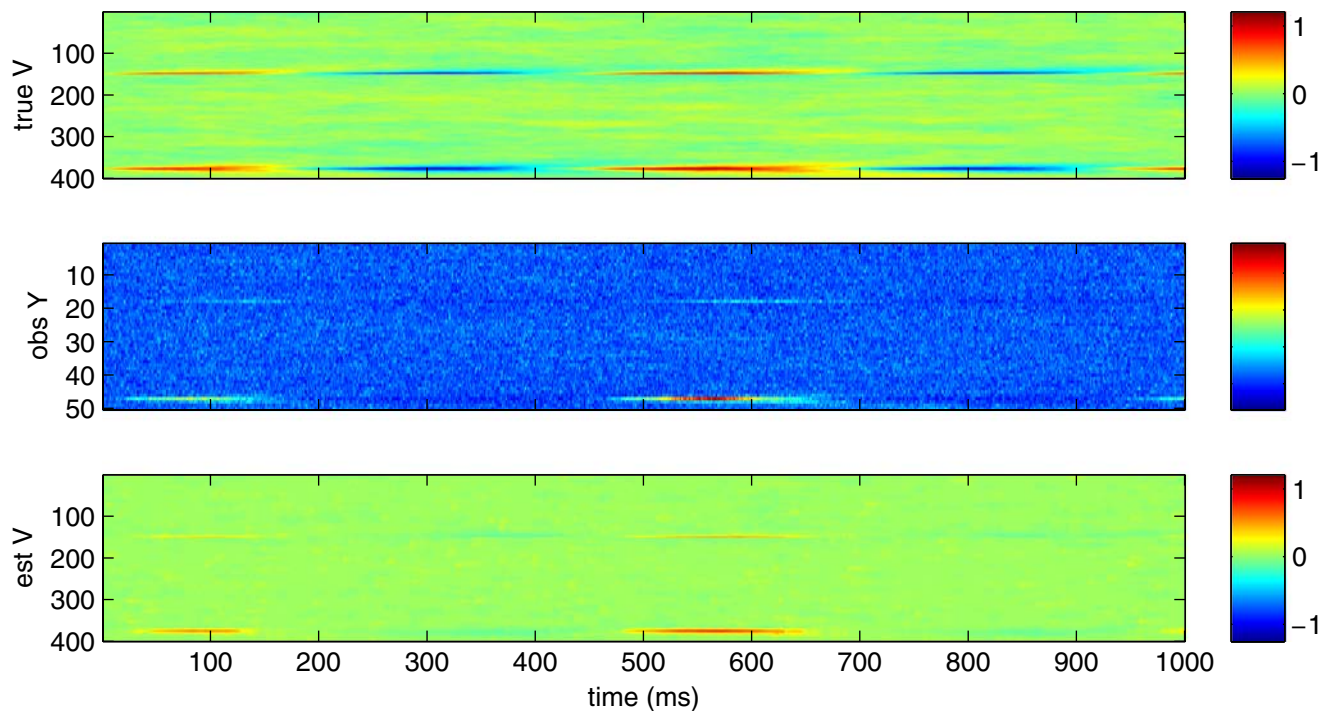


Fig. 11 Example of the approximate Kalman filter applied to nonlinear observations. *Top panel:* true voltage V . Sinusoidal current (2 Hz) was injected into a branch point of the tree (compartment 380); the tree was a simple one-branched structure with 400 compartments in this case. *Middle panel:* observations Y were generated by emulating the calcium-sensitive experiment described in the text. The voltage was spatially subsampled ($8x$),

transformed by a nonlinear activation curve $f(V)$, and then corrupted with noise. Note that most details about the subthreshold voltage are obscured by this nonlinear transformation $f(V)$. *Bottom panel:* voltage V recovered by the approximate Kalman filter. Note that superthreshold voltages are recovered fairly accurately, though subthreshold information is lost. The *top* and *bottom panels* share the same (rescaled) units

Note that initialization of this iteration is important here (unlike in the standard Kalman case), since the loglikelihood $\log p(y_t|V_t)$ is not necessarily concave in V_t . We have found that an effective strategy is to initialize V_{old} by taking one Kalman step forward from the point μ_{t-dt} , with $V_t(x)$ fixed at $E[V_t(x)|y_{xt}]$ (which may in turn be computed via a one-dimensional numerical integral if B_x is spatially localized); this corresponds to treating $E[V_t(x)|y_{xt}]$ as the observed data, with the noise variance set to zero. From this initialization we iterate to convergence, as described above. See Fig. 11 for an illustration.

5 Discussion

We have presented a collection of methods that make it possible to perform optimal quasi-linear filtering of noisy, subsampled voltage observations on large realistic dendritic trees. Perhaps the major current limitation of our approach is that we are forced to assume that the dynamics on the tree are constant in time, since we

make heavy use of the steady-state covariance C_0 , and this quantity is well-defined only in the case of time-invariant dynamics. It will be important to generalize this approach, perhaps via extended Kalman filter or hybrid particle filter methods (Doucet et al. 2001), to the case of active dendritic membranes (Johnston et al. 1996), where the dynamics depend strongly on the recent voltage history.

Another exciting potential direction for future research involves the optimal design of sampling schemes; we would like to design our imaging experiments so that each observation reduces our posterior uncertainty about the underlying voltage V (as summarized by the posterior state covariance C_t) as much as possible (Fedorov 1972; Lewi et al. 2009). As one example, it is intuitively plausible that observations of the voltage at or near branch points may be more informative than observations at the tips of dendrites (though we have not yet checked this intuition in numerical detail). Given a limited budget of compartments that may be sampled per unit of experimental time, it might be beneficial to optimize our experimental design accordingly,

by computing the expected reduction in the uncertainty C_t for each of a number of candidate sampling schemes, using the fast methods proposed here.

Our methods are based on low-rank approximations of the time-varying system covariance C_t (or more precisely, we represent C_t as a low-rank perturbation of the steady-state covariance C_0). A very promising related approach was recently described by Kellems et al. (2009); these authors used tools from the applied mathematical literature on low-rank approximations of high-dimensional linear systems (Antoulas 2005) to sharply reduce the dimensionality of quasi-active models used to describe subthreshold voltage dynamics. Our method is different: we do not truncate the system dynamics directly, but instead use the full dynamics to propagate a truncated approximation to the system covariance C_t . It will be interesting in the future to combine the benefits of these two approaches.

Acknowledgements LP is supported by a McKnight Scholar award and an NSF CAREER award. I thank Q. Huys, K. Rahnama Rad, and J. Vogelstein for many helpful conversations.

References

- Antoulas, A. (2005). *Approximation of large-scale dynamical systems*. Cambridge: Cambridge University Press.
- Araya, R., Eiselthal, K. B., & Yuste, R. (2006). Dendritic spines linearize the summation of excitatory potentials. *PNAS*, *103*(49), 18799–18804.
- Bell, J., & Craciun, G. (2005). A distributed parameter identification problem in neuronal cable theory models. *Mathematical Biosciences*, *194*(1), 1–19.
- Bloomfield, S., & Miller, R. (1986). A functional organization of ON and OFF pathways in the rabbit retina. *Journal of Neuroscience*, *6*(1), 1–13.
- Brette, R., Piwkowska, Z., Rudolph, M., Bal, T., & Destexhe, A. (2007). A nonparametric electrode model for intracellular recording. *Neurocomputing*, *70*, 1597–1601.
- Brockwell, P., & Davis, R. (1991). *Time Series: Theory and methods*. New York: Springer.
- Brown, E., Frank, L., Tang, D., Quirk, M., & Wilson, M. (1998). A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells. *Journal of Neuroscience*, *18*, 7411–7425.
- Canepari, M., Djuricic, M., & Zecevic, D. (2007). Dendritic signals from rat hippocampal CA1 pyramidal neurons during coincident pre- and post-synaptic activity: A combined voltage- and calcium-imaging study. *J Physiol*, *580*(2), 463–484.
- Chandrasekar, J., Kim, I., Bernstein, D., & Ridley, A. (2008). Cholesky-based reduced-rank square-root Kalman filtering. *American Control Conference*, *10*, 3987–3992.
- Coombes, S., Timofeeva, Y., Svensson, C. M., Lord, G. J., Josic, K., Cox, S. J., & Colbert, C. M. (2007). Branching dendrites with resonant membrane: a sum-over-trips approach. *Biological Cybernetics*, *97*(2), 137–149.
- Cox, S. (2004). Estimating the location and time course of synaptic input from multi-site potential recordings. *Journal of Computational Neuroscience*, *17*, 225–243.
- Cox, S., & Griffith, B. (2001). Recovering quasi-active properties of dendrites from dual potential recordings. *Journal of Computational Neuroscience*, *11*, 95–110.
- Cox, S. J., & Raol, J. H. (2004). Recovering the passive properties of tapered dendrites from single and dual potential recordings. *Mathematical Biosciences*, *190*(1), 9–37.
- Djuricic, M., Antic, S., Chen, W. R., & Zecevic, D. (2004). Voltage imaging from dendrites of mitral cells: EPSP attenuation and spike trigger zones. *Journal of Neuroscience*, *24*(30), 6703–6714.
- Djuricic, M., Popovic, M., Carnevale, N., & Zecevic, D. (2008). Functional structure of the mitral cell dendritic tuft in the rat olfactory bulb. *Journal of Neuroscience*, *28*(15), 4057–4068.
- Doucet, A., de Freitas, N., & Gordon, N., (Eds.) (2001). *Sequential Monte Carlo in Practice*. New York: Springer.
- Durbin, J., & Koopman, S. (2001). *Time series analysis by state space methods*. Oxford: Oxford University Press.
- Fahrmeir, L., & Tutz, G. (1994). *Multivariate statistical modelling based on generalized linear models*. New York: Springer.
- Fedorov, V. (1972). *Theory of optimal experiments*. New York: Academic.
- Gilljins, S., Bernstein, D., & De Moor, B. (2006). The reduced rank transform square root filter for data assimilation. In *Proc. of the 14th IFAC Symposium on System Identification*.
- Gobel, W., & Helmchen, F. (2007). New angles on neuronal dendrites *in vivo*. *Journal of Neurophysiology*, *98*(6), 3770–3779.
- Hines, M. (1984). Efficient computation of branched nerve equations. *International Journal of Bio-Medical Computing*, *15*(1), 69–76.
- Holekamp, T., Turaga, D., & Holy, T. (2008). Fast three-dimensional fluorescence imaging of activity in neural populations by objective-coupled planar illumination microscopy. *Neuron*, *57*, 661–672.
- Howard, A., & Jebara, T. (2005). Square root propagation. *Columbia University Computer Science Technical Reports*, 040-05.
- Huys, Q., Ahrens, M., & Paninski, L. (2006). Efficient estimation of detailed single-neuron models. *Journal of Neurophysiology*, *96*, 872–890.
- Huys, Q., & Paninski, L. (2009). Model-based smoothing of, and parameter estimation from, noisy biophysical recordings. *PLOS Computational Biology*, *5*, e1000379.
- Johnston, D., Magee, J. C., Colbert, C. M., & Christie, B. R. (1996). Active properties of neuronal dendrites. *Annual review Neuroscience*, *19*, 165–186.
- Jordan, M. I. (Ed.) (1999). *Learning in graphical models*. Cambridge: MIT.
- Kellems, A., Roos, D., Xiao, N., & Cox, S. (2009). Low-dimensional, morphologically accurate models of subthreshold membrane potential. *Journal of Computational Neuroscience*, *27*, 161–176.
- Knopfel, T., Diez-Garcia, J., & Akemann, W. (2006). Optical probing of neuronal circuit dynamics: genetically encoded versus classical fluorescent sensors. *Trends in Neurosciences*, *29*, 160–166.
- Koch, C. (1984). Cable theory in neurons with active, linearized membranes. *Biological Cybernetics*, *50*, 15–33.
- Kole, M., Ischner, S., Kampa, B., Williams, S., Ruben, P., & Stuart, G. (2008). Action potential generation requires a

- high sodium channel density in the axon initial segment. *Nature Neuroscience*, *11*, 178–186.
- Larkum, M. E., Watanabe, S., Lasser-Ross, N., Rhodes, P., & Ross, W. N. (2008). Dendritic properties of turtle pyramidal neurons. *Journal of neurophysiology*, *99*(2), 683–694.
- Lewi, J., Butera, R., & Paninski, L. (2009). Sequential optimal design of neurophysiology experiments. *Neural Computation*, *21*, 619–687.
- Manwani, A., & Koch, C. (1999). Detecting and Estimating Signals in Noisy Cable Structures, I: Neuronal Noise Sources. *Neural Computation*, *11*(8), 1797–1829.
- Morse, T., Davison, A., & Hines, M. (2001). Parameter space reduction in neuron model optimization through minimization of residual voltage clamp current. *Society for Neuroscience Abstracts*.
- Palmer, L. M., & Stuart, G. J. (2006). Site of action potential initiation in layer 5 pyramidal neurons. *Journal of Neuroscience*, *26*(6), 1854–1863.
- Paninski, L., Ahmadian, Y., Ferreira, D., Koyama, S., Rahnama, K., Vidne, M., et al. (2009). A new look at state-space models for neural data. doi:10.1007/s10827-009-0179-x.
- Paninski, L., & Ferreira, D. (2008). State-space methods for inferring synaptic inputs and weights. COSYNE.
- Petrusca, D., Grivich, M. I., Sher, A., Field, G. D., Gauthier, J. L., Greschner, M., et al. (2007). Identification and characterization of a Y-like primate retinal ganglion cell type. *Journal of Neuroscience*, *27*(41), 11019–11027.
- Press, W., Teukolsky, S., Vetterling, W., & Flannery, B. (1992). *Numerical recipes in C*. Cambridge: Cambridge University Press.
- Sabino, J. (2007). *Solution of large-scale Lyapunov equations via the block modified Smith method*. PhD thesis, Rice University.
- Sacconi, L., Dombeck, D. A., & Webb, W. W. (2006). Overcoming photodamage in second-harmonic generation microscopy: Real-time optical recording of neuronal action potentials. *Proceedings of the National Academy of Sciences*, *103*(9), 3124–3129.
- Shental, O., Bickson, D., Siegel, P. H., Wolf, J. K., & Dolev, D. (2008). Gaussian belief propagation for solving systems of linear equations: Theory and application. [arXiv:0810.1119v1](https://arxiv.org/abs/0810.1119v1).
- Sjostrom, P. J., Rancz, E. A., Roth, A., & Hausser, M. (2008). Dendritic Excitability and Synaptic Plasticity. *Physiological Reviews*, *88*(2), 769–840.
- Spruston, N. (2008). Pyramidal neurons: dendritic structure and synaptic integration. *Nature Reviews Neuroscience*, *9*, 206–221.
- Stuart, G., & Sakmann, B. (1994). Active propagation of somatic action potential into neocortical pyramidal cell dendrites. *Nature*, *367*, 69–72.
- Stuart, G., Spruston, N., & Häusser, M., (Eds.) (1999). *Dendrites*. Oxford: Oxford University Press.
- Treebushny, D., & Madsen, H. (2005). On the construction of a reduced rank square-root kalman filter for efficient uncertainty propagation. *Future Generations Computer Systems*, *21*, 1047–1055.
- Verlaan, M. (1998). *Efficient Kalman filtering algorithms for hydrodynamic models*. PhD thesis, TU Delft.
- Vucinic, D., & Sejnowski, T. J. (2007). A compact multiphoton 3d imaging system for recording fast neuronal activity. *PLoS ONE*, *2*(8), e699.
- Weiss, Y., & Freeman, W. T. (2001). Correctness of belief propagation in gaussian graphical models of arbitrary topology. *Neural Computation*, *13*(10), 2173–2200.
- Wood, R., Gurney, K., & Wilson, C. (2004). A novel parameter optimisation technique for compartmental models applied to a model of a striatal medium spiny neuron. *Neurocomputing*, *58–60*, 1109–1116.