

A new look at state-space models for neural data

Liam Paninski

Department of Statistics and Center for Theoretical Neuroscience
Columbia University

<http://www.stat.columbia.edu/~liam>

liam@stat.columbia.edu

May 16, 2010

Support: CRCNS, Sloan Fellowship, NSF CAREER, McKnight Scholar award.

State-space models

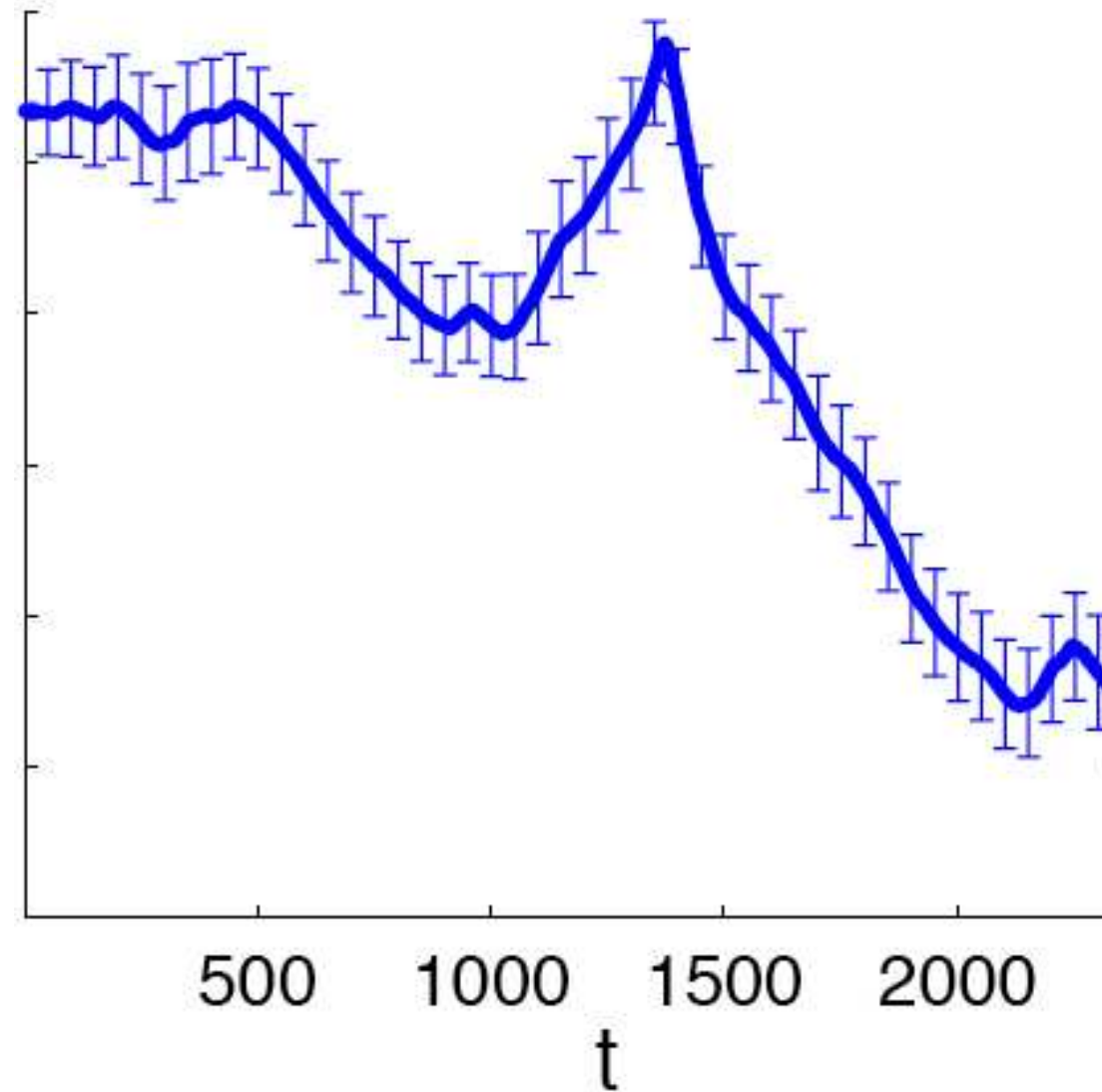
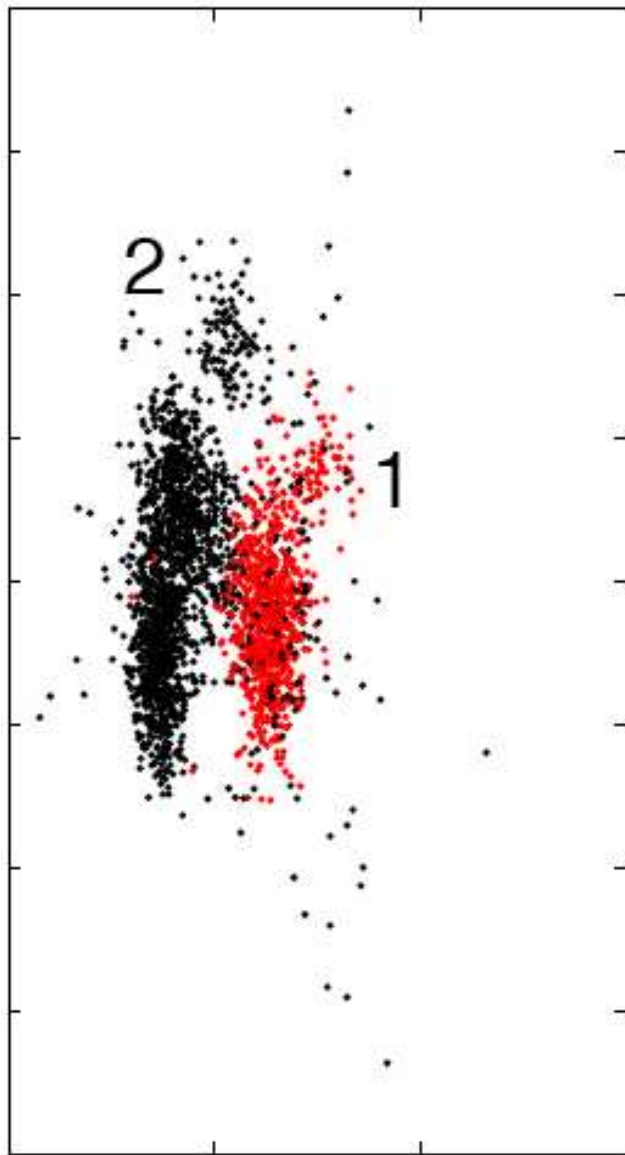
Unobserved state q_t with Markov dynamics $p(q_{t+1}|q_t)$ (i.e., q_t is a noisy dynamical system)

Observed y_t : $p(y_t|q_t)$ (noisy, partial observations)

Goal: infer $p(q_t|Y_{0:T})$

Dozens of applications in neuroscience (Paninski et al., 2010).

Example: nonstationary spike sorting



q_t : mean waveform; y_t : observed spikes (Calabrese '10); data from (Harris et al., 2000)

Basic paradigm: the forward recursion

We want $p(q_t|Y_{1:t}) \propto p(q_t, Y_{1:t})$. We know that

$$p(Q, Y) = p(Q)p(Y|Q) = p(q_1) \left(\prod_{t=2}^T p(q_t|q_{t-1}) \right) \left(\prod_{t=1}^T p(y_t|q_t) \right)$$

To compute $p(q_t, Y_{1:t})$ recursively, just write out marginal and pull out constants from the integrals:

$$\begin{aligned} p(q_t, Y_{1:t}) &= \int_{q_1} \int_{q_2} \cdots \int_{q_{t-1}} p(Q_{1:t}, Y_{1:t}) = \int_{q_1} \int_{q_2} \cdots \int_{q_{t-1}} p(q_1) \left(\prod_{i=2}^t p(q_i|q_{i-1}) \right) \left(\prod_{i=1}^t p(y_i|q_i) \right) \\ &= p(y_t|q_t) \int_{q_{t-1}} p(q_t|q_{t-1})p(y_{t-1}|q_{t-1}) \int_{q_{t-2}} \cdots \int_{q_2} p(q_3|q_2)p(y_2|q_2) \int_{q_1} p(q_2|q_1)p(y_1|q_1)p(q_1). \end{aligned}$$

So, just recurse

$$p(q_t, Y_{1:t}) = p(y_t|q_t) \int_{q_{t-1}} p(q_t|q_{t-1})p(q_{t-1}, Y_{1:t-1}).$$

Linear-Gaussian (Kalman) case: requires $O(\dim(q)^3 T)$ time; just matrix algebra.

Approximate solutions in more general case, e.g., Gaussian approximations (Brown et al., 1998), or Monte Carlo (“particle filtering”).

Key point: efficient recursive computations $\implies O(T)$ time.

Computing the MAP path

We often want to compute the MAP estimate

$$\hat{Q} = \arg \max_Q p(Q|Y).$$

In standard Kalman setting, forward-backward gives MAP (because $E(Q|Y)$ and \hat{Q} coincide in Gaussian case).

More generally, extended Kalman-based methods give approximate MAP, but are non-robust: forward distribution $p(q_t|Y_{0:t})$ may be highly non-Gaussian even if full joint distribution $p(Q|Y)$ is nice and unimodal.

Write out the posterior:

$$\begin{aligned}\log p(Q|Y) &= \log p(Q) + \log p(Y|Q) \\ &= \sum_t \log p(q_{t+1}|q_t) + \sum_t \log p(y_t|q_t)\end{aligned}$$

Two basic observations:

- If $\log p(q_{t+1}|q_t)$ and $\log p(y_t|q_t)$ are concave, then so is $\log p(Q|Y)$.
- Hessian H of $\log p(Q|Y)$ is block-tridiagonal: $p(y_t|q_t)$ contributes a block-diag term, and $\log p(q_{t+1}|q_t)$ contributes a block-tridiag term.

Now recall Newton's method: iteratively solve $HQ_{dir} = \nabla$. Solving tridiagonal systems requires $O(T)$ time.

— computing MAP by Newton's method requires $O(T)$ time, even in highly non-Gaussian cases.

Newton here acts as an iteratively reweighted Kalman smoother (Fahrmeir and Kaufmann, 1991; Davis and Rodriguez-Yam, 2005; Jungbacker and Koopman, 2007); all suff. stats may be obtained in $O(T)$ time.

Constrained optimization

In many cases we need to impose constraints (e.g., nonnegativity) on q_t . Easy to incorporate here, via interior-point (barrier) methods:

$$\begin{aligned}\arg \max_{Q \in C} \log p(Q|Y) &= \lim_{\epsilon \searrow 0} \arg \max_Q \left\{ \log p(Q|Y) + \epsilon \sum_t f(q_t) \right\} \\ &= \lim_{\epsilon \searrow 0} \arg \max_Q \left\{ \sum_t \log p(q_{t+1}|q_t) + \log p(y_t|q_t) + \epsilon f(q_t) \right\};\end{aligned}$$

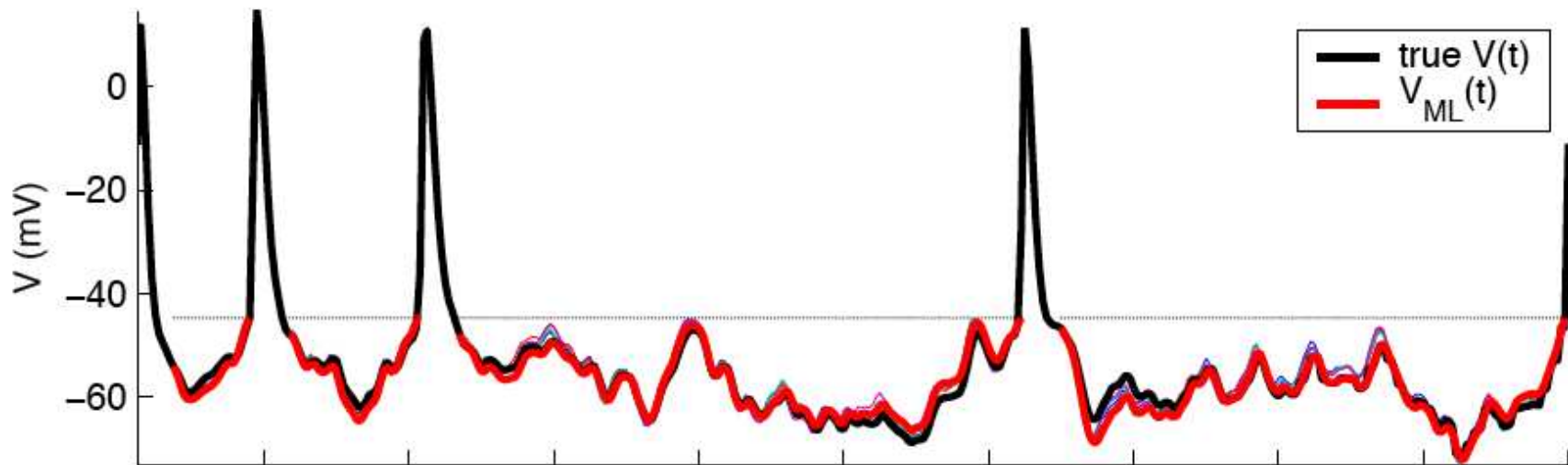
$f(\cdot)$ is concave and approaching $-\infty$ near boundary of constraint set C . The Hessian remains block-tridiagonal and negative semidefinite for all $\epsilon > 0$, so optimization still requires just $O(T)$ time.

Example: computing the MAP subthreshold voltage given superthreshold spikes

Leaky, noisy integrate-and-fire model:

$$V(t + dt) = V(t) - dtV(t)/\tau + \sigma\sqrt{dt}\epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, 1)$$

Observations: $y_t = 0$ (no spike) if $V_t < V_{th}$; $y_t = 1$ if $V_t = V_{th}$

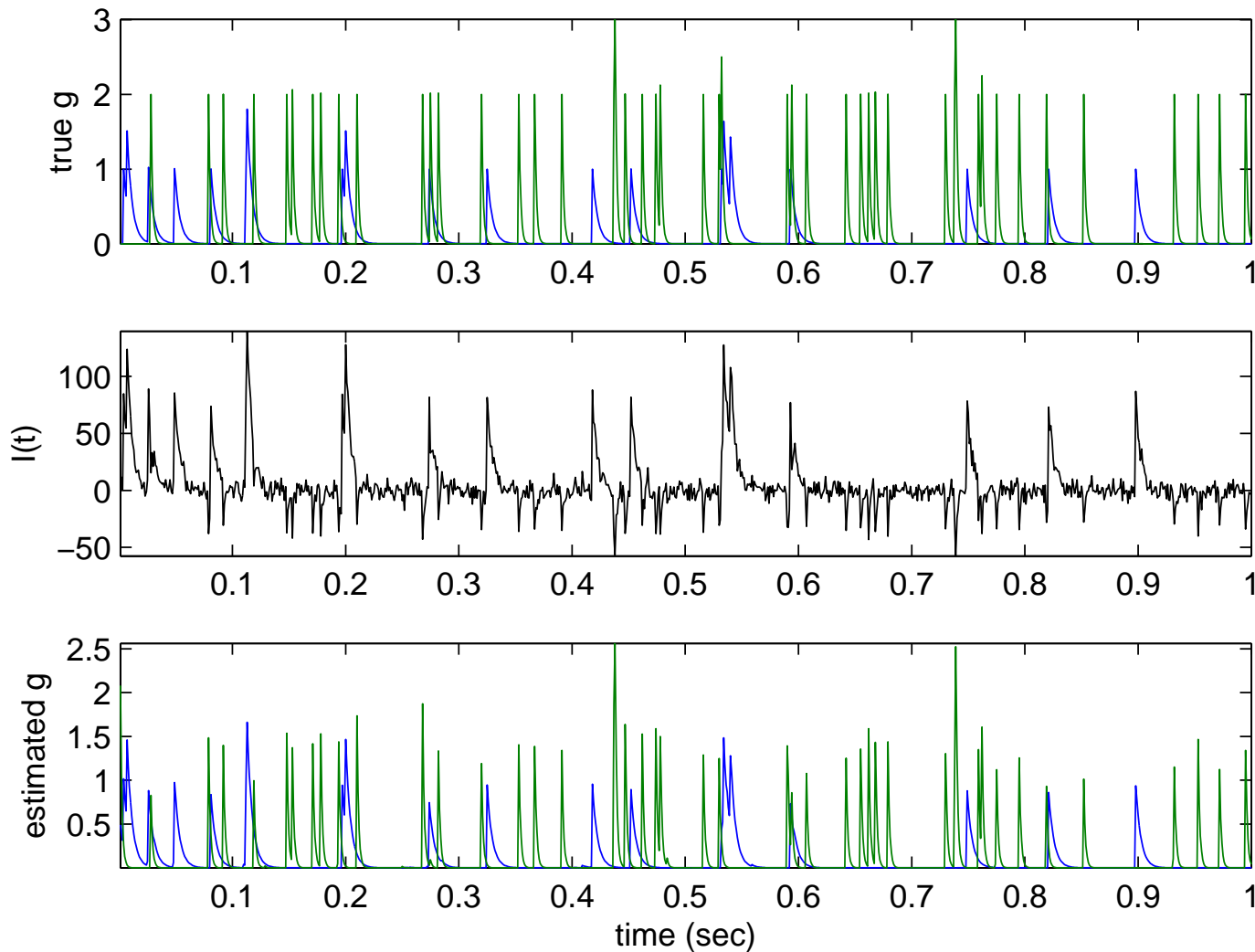


(Paninski, 2006)

Example: inferring presynaptic input

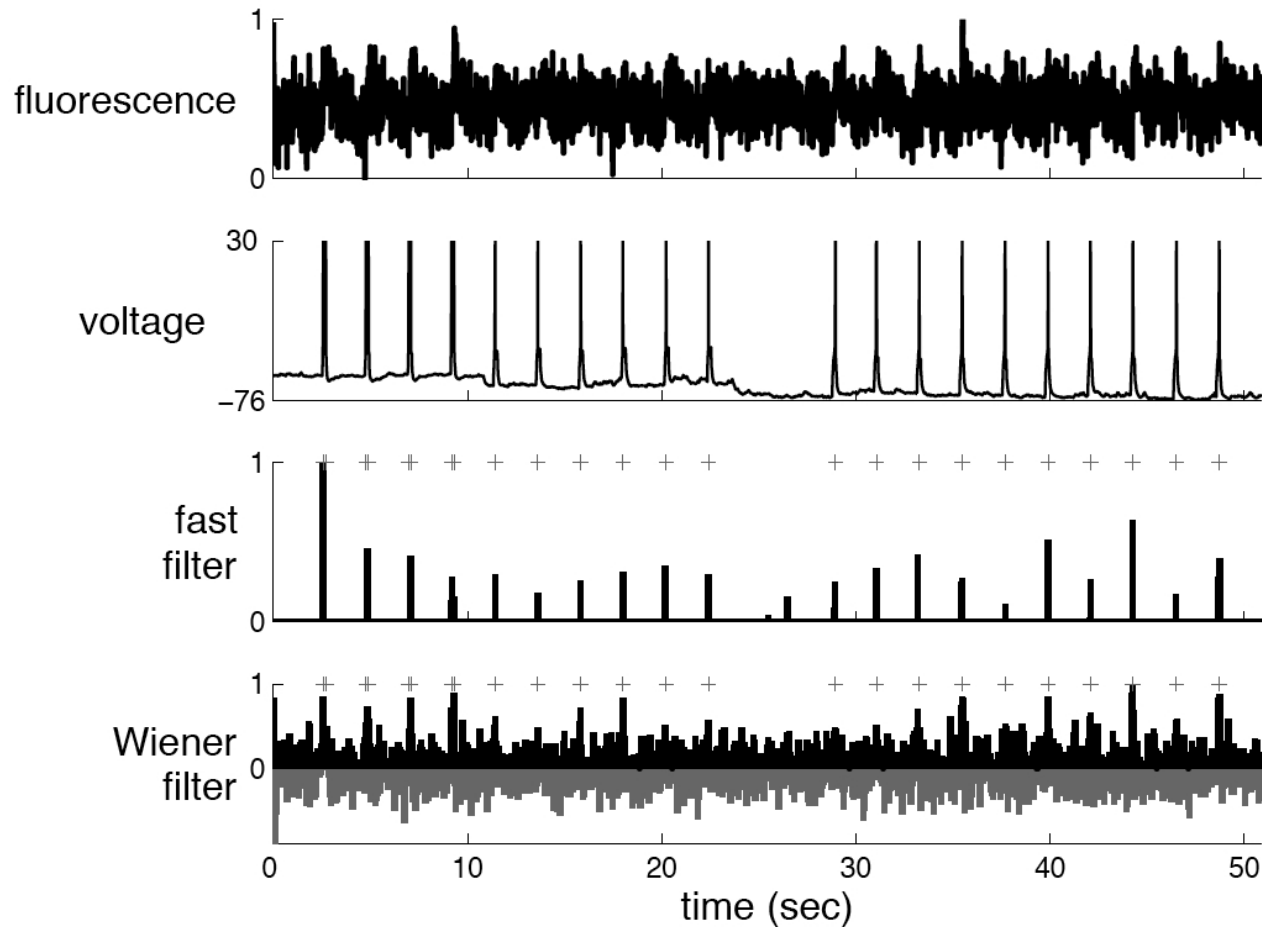
$$I_t = \sum_j g_j(t)(V_j - V_t)$$

$$g_j(t + dt) = g_j(t) - dtg_j(t)/\tau_j + N_j(t), \quad N_j(t) > 0$$



Example: inferring spike times from slow, noisy calcium data

$$C(t + dt) = C(t) - dtC(t)/\tau + N_t; \quad N_t > 0; \quad y_t = C_t + \epsilon_t$$



— nonnegative deconvolution is a recurring problem in signal processing; many other possible applications (Vogelstein et al., 2008).

Optimal control of spike timing

Optimal experimental design and neural prosthetics applications require us to perturb the network at will. How can we make a neuron fire exactly when we want it to?

Assume bounded inputs; otherwise problem is trivial.

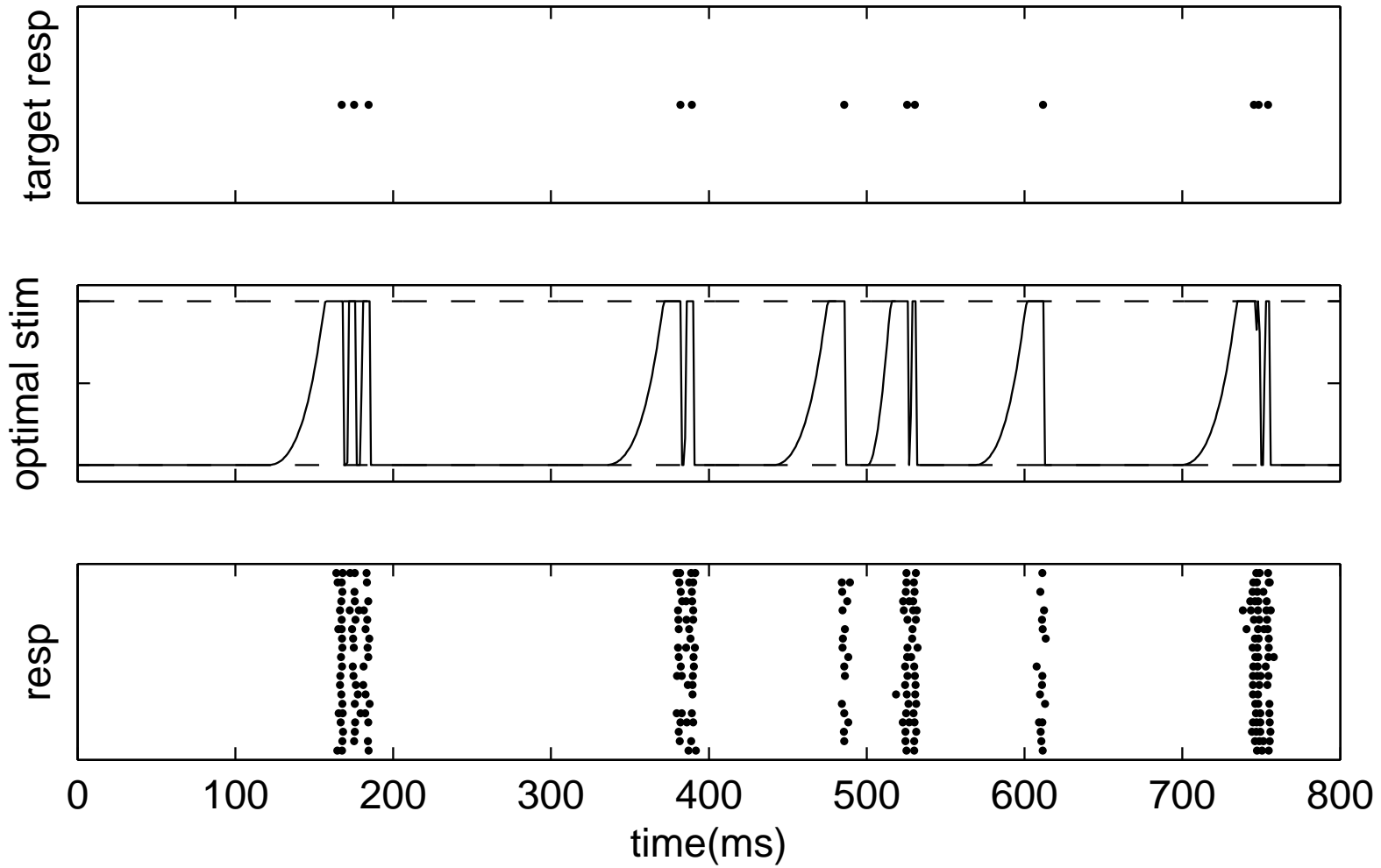
Start with a simple model:

$$\lambda_t = f(\vec{k} * I_t + h_t).$$

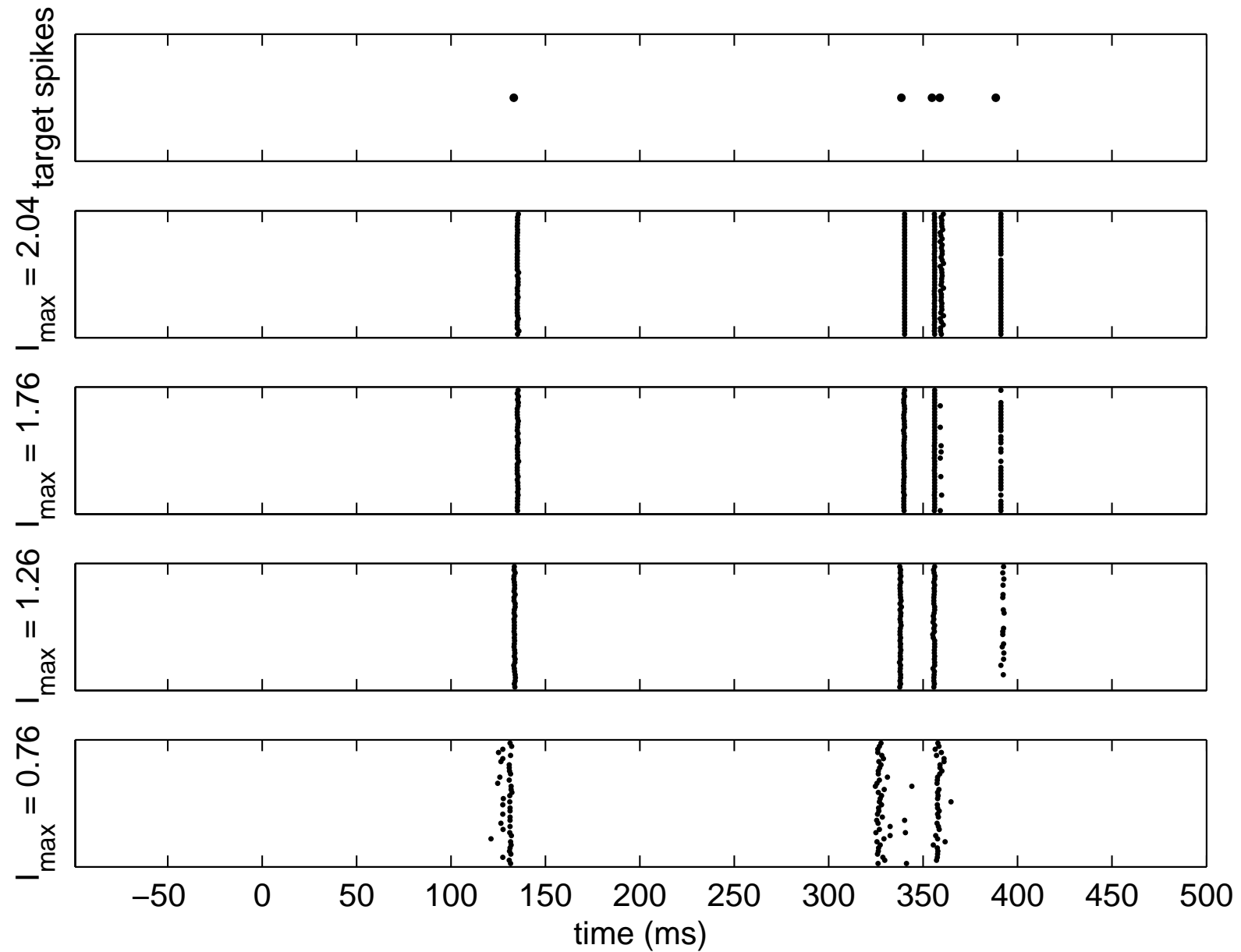
Now we can just optimize the likelihood of the desired spike train, as a function of the input I_t , with I_t bounded.

Concave objective function over convex set of possible inputs I_t
+ Hessian is banded $\implies O(T)$ optimization.

Optimal electrical control of spike timing



Example: intracellular control of spike timing

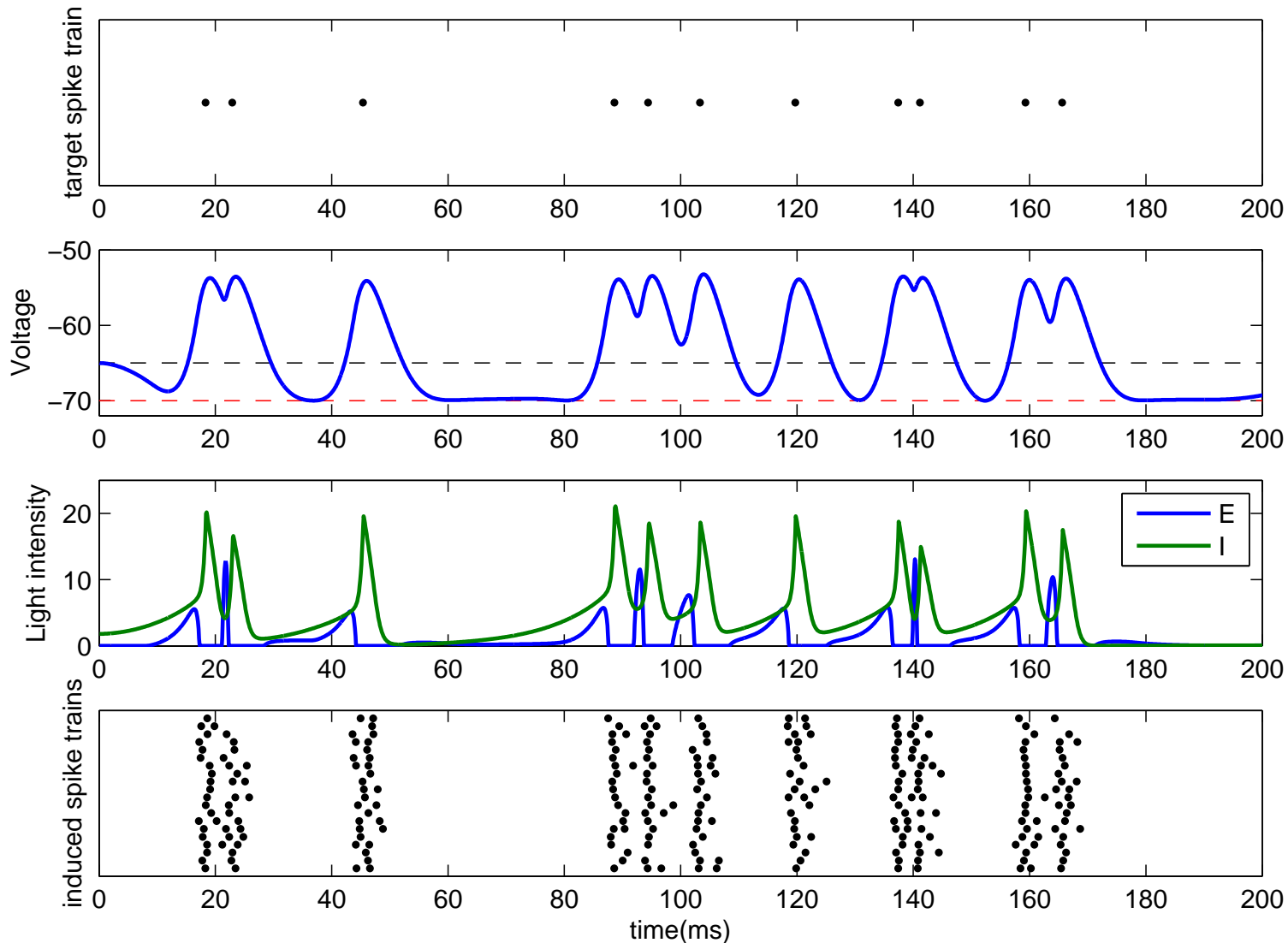


(Ahmadian et al 2010)

Optical conductance-based control of spiking

$$V_{t+dt} = V_t + dt \left(-gV_t + g_t^i(V^i - V_t) + g_t^e(V^e - V_t) \right) + \sqrt{dt}\sigma\epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0,1)$$

$$g_{t+dt}^i = g_t^i + dt \left(-\frac{g_t^i}{\tau_i} + a_{ii}L_t^i + a_{ie}L_t^e \right); \quad g_{t+dt}^e = g_t^e + dt \left(-\frac{g_t^e}{\tau_e} + a_{ee}L_t^e + a_{ei}L_t^i \right)$$



Conclusions

- GLM and state-space approaches provide flexible, powerful methods for answering key questions in neuroscience
- Close relationships between forward-backward methods familiar from state-space theory and banded matrices familiar from spline theory
- Log-concavity, banded matrix methods make computations very tractable

References

- Brown, E., Frank, L., Tang, D., Quirk, M., and 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.
- Davis, R. and Rodriguez-Yam, G. (2005). Estimation for state-space models: an approximate likelihood approach. *Statistica Sinica*, 15:381–406.
- Fahrmeir, L. and Kaufmann, H. (1991). On Kalman filtering, posterior mode estimation and fisher scoring in dynamic exponential family regression. *Metrika*, 38:37–60.
- Harris, K., Henze, D., Csicsvari, J., Hirase, H., and Buzsaki, G. (2000). Accuracy of tetrode spike separation as determined by simultaneous intracellular and extracellular measurements. *J. Neurophys.*, 84:401–414.
- Jungbacker, B. and Koopman, S. (2007). Monte Carlo estimation for nonlinear non-Gaussian state space models. *Biometrika*, 94:827–839.
- Paninski, L. (2006). The most likely voltage path and large deviations approximations for integrate-and-fire neurons. *Journal of Computational Neuroscience*, 21:71–87.
- Paninski, L., Ahmadian, Y., Ferreira, D., Koyama, S., Rahnama, K., Vidne, M., Vogelstein, J., and Wu, W. (2010). A new look at state-space models for neural data. *Journal of Computational Neuroscience*, In press.
- Vogelstein, J., Babadi, B., Watson, B., Yuste, R., and Paninski, L. (2008). Fast nonnegative deconvolution via tridiagonal interior-point methods, applied to calcium fluorescence data. *Statistical analysis of neural data (SAND) conference*.