I-16

Optimal experimental design for sampling voltage on dendritic trees in the low-SNR regime

Jonathan Huggins[1], Liam Paninski[1]

[1]: Columbia University

Introduction

We use a Kalman filtering framework to develop optimal experimental design methods for voltage sampling. Our approach is to use a simple greedy algorithm with lazy evaluation to minimize the expected square error of the estimated spatiotemporal voltage signal. We take advantage of some particular features of the dendritic filtering problem to efficiently calculate the Kalman estimator's covariance.

Background: Voltage Sensing

- Underlying experimental task: voltage sensing
- State-of-the-art, random access, laser-based techniques are low-SNR
- But measurements are sparse
- If we have the complete spatiotemporal signal, can calculate biophysical quantities of interest

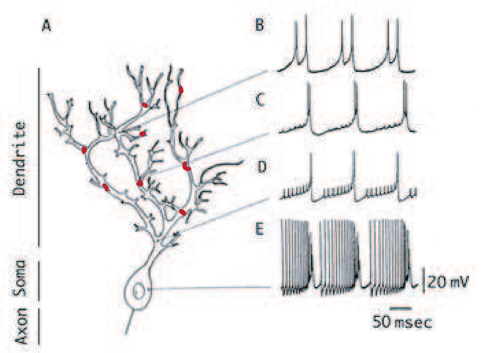


Figure: Modified from Llinas and Sugimori 1980

- Need to infer voltage across the tree
- Would like to choose the best sampling sites on the tree

A linear-Gaussian model for voltage and observation dynamics

To model the dendritic dynamics, we use the standard cable equation:

$$V_{t+dt}(x) = V_t(x) + dt(-g_x V_{t+dt}(x) + \sum_{w \in N(x)} a_{xw}[V_{t+dt}(w) - V_{t+dt}(x)]), \quad (1)$$

where $V_t(x)$ denotes the voltage in compartment x at time t, g_x is the membrane conductance, $N(x)$ is the set of adjoining compartments, and a_{xw} is the intercompartmental conductance. Eq. (1) can be written in matrix form as

$$V_{d+dt} = AV_{dt},$$

where $A = (I - K)^{-1}$ and K is a matrix in "Hines" form [Hines, 1984]. From here we can write out the the dynamics and observation equations as

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

$$y_t = B_t V_t + \eta_t, \; \eta_t \sim \mathcal{N}(0, W_t), \quad (3)$$

where $\{y_t\}$ are the vectors of the observed voltages, B_t is a matrix that specifies how the observations are related instantaneously to the voltage vector, and W_t is the covariance matrix that defines the noisiness of the observations. The forward covariance matrix $C_t^f = Cov(V_t|Y_{1:t})$ can be written using the Kalman recursion as

$$C_t^f = \left[(AC_{t-dt}^f A^T + \sigma^2 dtI)^{-1} + B_t^T W_t^{-1} B_t\right]^{-1}. \quad (4)$$

It follows from Eq. (4) that the steady state covariance matrix C_0 is

$$C_0 = AC_0A^T + \sigma^2 dtI = \sigma^2 dt(I - A^2)^{-1}. \quad (5)$$

In the limit as $dt \to 0$,

$$C_0 \to -\frac{1}{2}K^{-1}, \quad (6)$$

where K^{-1} can be interpreted as the transfer impedance matrix.

Measuring sampling scheme quality with covariance-based metrics

- Can use the smoothed covariances
$$C_t^s = Cov(V_t|Y_{1:T}) \quad (7)$$
to measure the quality of a sampling scheme
- C_t^s does not depend on the data
- Objective function: (weighted) mean squared error (MSE) = (weighted) summed variance:
$$v_w(\mathcal{O}) = \sum_{t=0}^T \sum_{i=1}^N w(i,t)[C_t^s]_{ii} \quad (8)$$

The optimal MSE can be computed by a fast Kalman recursion in $\mathcal{O}(NT)$ time

- Standard methods to calculate mean and covariances require $\mathcal{O}(N^3)$ time and $\mathcal{O}(N^2)$ space, which is not practical in the case of $N \sim 10^4$
- [Paninski, 2010] showed how to calculate C_t^f in $\mathcal{O}(N)$ time and space using a low rank approximation to C_t^f:
$$C_t^f \approx C_0 + U_t D_t U_t^T. \quad (9)$$
- Similar techniques allow us to approximate the smoothed covariance by an expression of the same form
- Intuitively, the low rank approximation (9) is justified as follows: if we make k observations at $t = 1$ then, (9) holds exactly with U_1 having rank k. If we make no further observations, then C_t^f follows the update rule
$$C_t^f = C_0 + AU_{t-1}D_{t-1}U_{t-1}U_{t-1}A^T \quad (10)$$
Iterating the equation gives
$$C_t^f = C_0 + A^{t-1}U_1D_1U_1^T(A^{t-1})^T. \quad (11)$$
The second term will decay exponentially; for t sufficiently large, we can discard some dimensions of the perturbation $AU_{t-1}D_{t-1}U_{t-1}U_{t-1}A^T$ without experiencing much error in C_t^f.

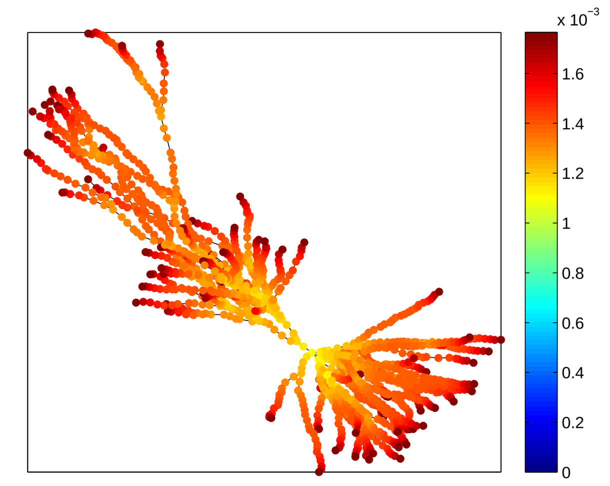
Using a greedy algorithm with lazy evaluation to choose an optimal sampling scheme

- In general choosing the best k sample locations is an NP-hard problem
- If the objective function is submodular, then it can be efficiently optimized via the greedy algorithm [Nemhauser et al., 1978, Krause et al., 2008]
- Submodularity is an intuitive diminishing returns property: the more observations added, the smaller the increase achieved by each additional observation.
- While $v(\cdot)$ is not generally submodular, in many cases an equivalent function, the variance reduction, is submodular [Das and Kempe, 2008]. Define the variance reduction $\rho(\cdot)$ as
$$\rho(\mathcal{O}) := v(\emptyset) - v(\mathcal{O}) = -\sum_{t=0}^T \mathrm{tr}(P_t G_t P_t^T), \quad (12)$$
- Intuitively, we would expect $\rho(\cdot)$ to be nearly submodular in this case because, as more observations are added, additional observations will contribute smaller amounts of new information. Thus, these new observations will result in smaller decreases in the variance.
- Empirically $\rho(\cdot)$ proved to be almost submodular
- Lazy evaluation allows the algorithm to only re-evaluate the objective function at a few compartments per iteration, which led to about a three orders of magnitude time savings

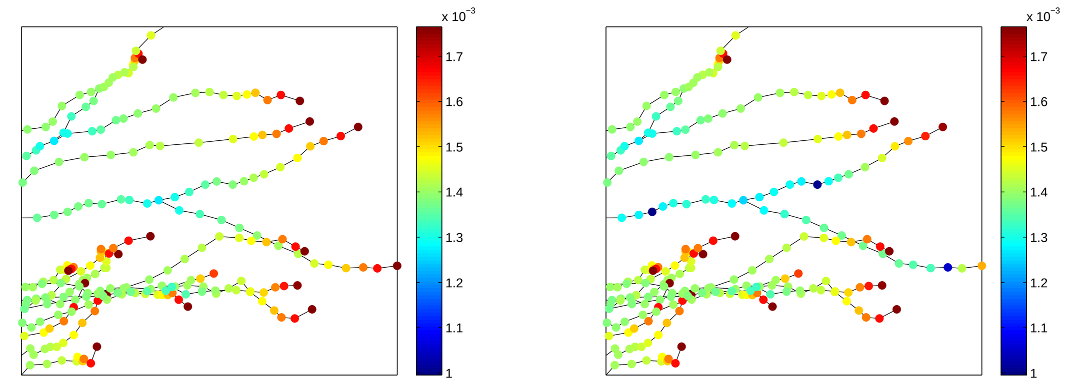
Computation time remains a issue for practical usage

- The magnitude of the computation time τ, the number of observations k proved to be problematic for large k
- In the time-invariant case, for $k = 10, \tau \approx 10$ minutes, for $k = 30, \tau \approx 1.5$ hours, and for $k = 100, \tau$ jumped to almost 2 days
- In time-varying case, for $k = 10, \tau \approx 30$ minutes, but jumped to 20 hours for k = 20
- The time variant implementation probably remains impractical without either an efficient parallelized implementation or spatial downsampling

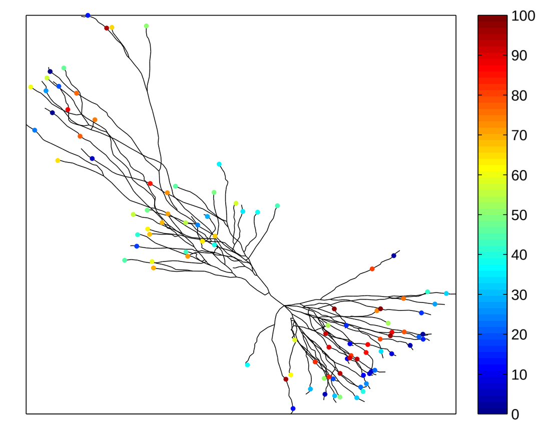
The optimal method samples from compartments near where the steady state covariance is largest



- *Above*: Relative magnitudes of the prior variances of a pyramidal cell geometry
- The variance increases farther away from the soma
- Observing at or near the tips, therefore, has the potential to provide the largest total reduction

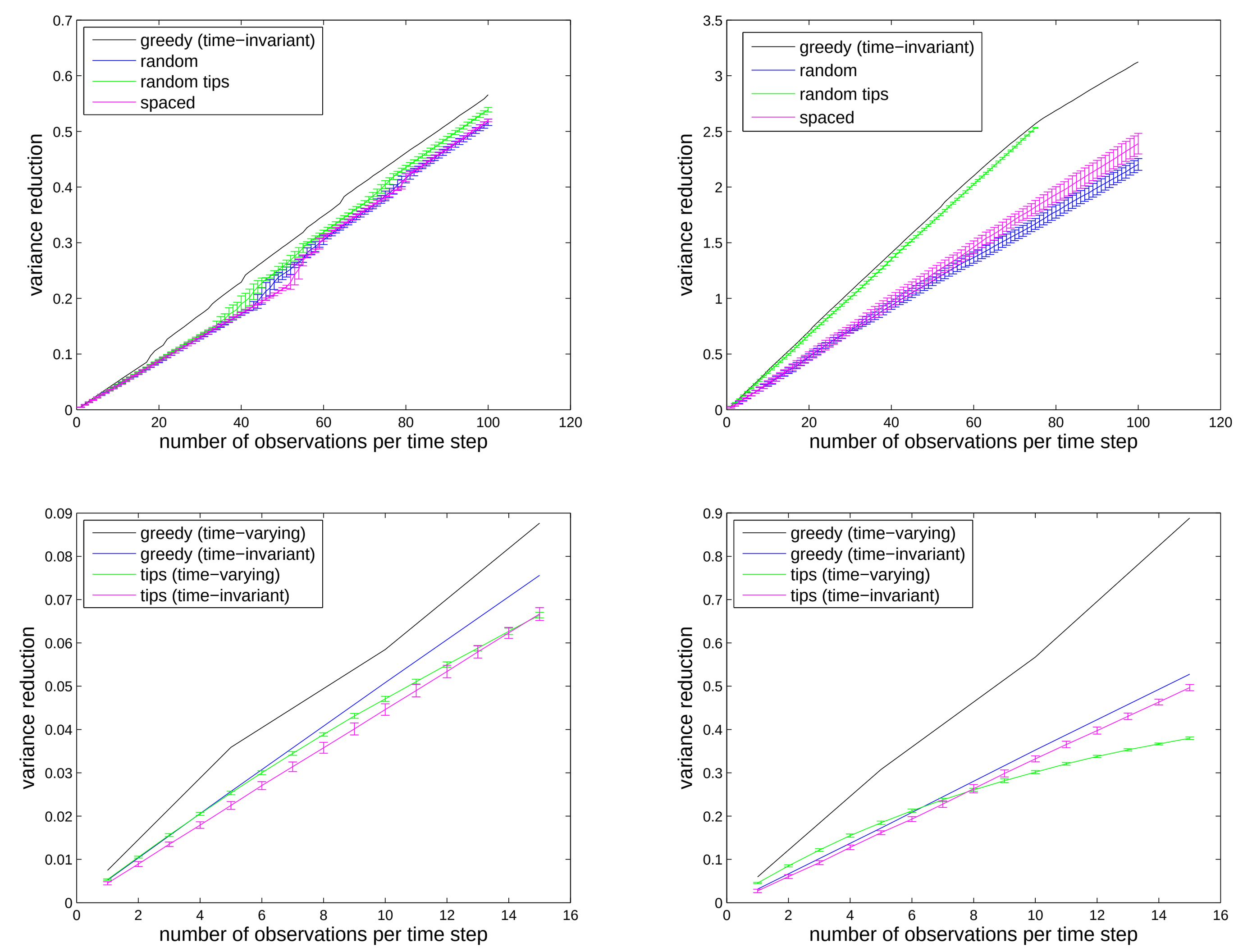


- *Right*: A comparison of the variances of the compartments of a subtree of the pyramidal cell before and after making three three observations
- The variance reduction effect is strongest around the observation close to a dendritic tip



- *Left*: Sampling scheme generated by greedily selecting 100 observation locations
- The colors indicate the order the locations were selected by the greedy algorithm.
- As expected, the greedy algorithm heavily favors sampling at locations near dendritic tips

The optimal method outperforms simpler heuristics in the case of time-variant sampling



Conclusions

- Low-rank perturbation methods allow for efficient computation of the smoothed covariance, which can be used to calculate a number of measures of experimental optimality
- We have shown how to tractably design an optimal sampling scheme using one possible metric
- In the simplest case of spatially-constant noise and variance weighting, the optimal greedy algorithm can be well-approximated by simpler heuristics
- For time varying sampling schemes, the greedy algorithm outperformed the simpler methods, however
- High computational requirements still remain a problem

Future Work

- Generalize results to non-linear and spiking cases
- Generalize to allow time-variant dynamics
- Investigate other metrics such as mutual information

Acknowledgments

JHH is supported by the Columbia College Rabi Scholars Program. LP is supported by a McKnight Scholar award and an NSF CAREER award. Computation performed on the Columbia Hotfoot HPC Research Cluster. We thank K. Rahnama Rad and C. Guestrin for helpful conversations.

References

A. Das and D. Kempe. Algorithms for subset selection in linear regression. In *Proceedings of the 40th annual ACM symposium on Theory of computing*, STOC '08, pages 45–54, 2008.

M. Hines. Efficient computation of branched nerve equations. *International Journal of Bio-Medical Computing*, 15(1):69–76, 1984.

A. Krause, A. Singh, and C. Guestrin. Near-optimal sensor placements in gaussian processes: Theory, efficient algorithms and empirical studies. *Journal of Machine Learning Research*, 9: 235–284, 2008.

G. Nemhauser, L. Wolsey, and J. Fisher. An analysis of the approximations for maximizing submodular set functions. *Mathematical Programming*, 14:265–294, 1978.

L. Paninski. Fast kalman filtering on quasilinear dendritic trees. *Journal of Computational Neuroscience*, 28:211–28, 2010.

email: jhh2143@columbia.edu, liam@stat.columbia.edu paper available at http://stat.columbia.edu/∼liam/research/pubs/huggins-opt-design.pdf