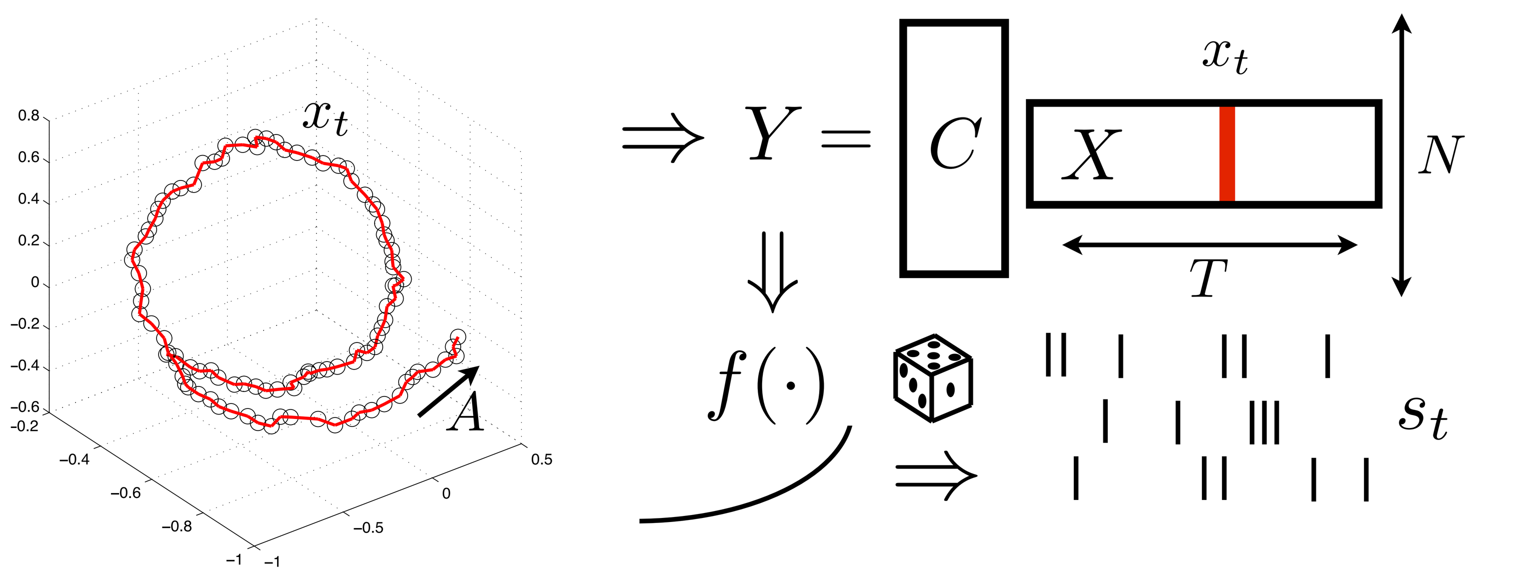
Robust Learning of Dynamics for Large Neural Ensembles

David Pfau, Eftychios Pnevmatikakis, Liam Paninski
{pfau@neurotheory, eftychios@stat, liam@stat}.columbia.edu
Columbia University

Overview

Recordings from large populations of neurons are increasingly common. How can we reduce the number of dimensions of our neural signal without throwing away relevant information? Can we recover models of the dynamics of neural signals in a way that is robust to outliers, nonstationarity and other deviations from a model? Can we separate global dynamics from local connections?

Linear Dynamical GLM



Dynamics and input: $x_{t+1} = Ax_t + Bu_t + \epsilon_t$

Pre-nonlinearity rate: $y_t = Cx_t + Ds_{t-k:t-1}$

Spikes: $s_t \sim \xi(f(y_t))$

Linear (but not necessarily Gaussian) dynamics, with lower dimension than number of spikes, followed by linear-nonlinear-point-process (Poisson, Bernoulli...) spiking, possibly including dependence on spike history. (Input and spike history terms are not included in the diagram)

Existing learning methods: Expectation-Maximization [5,7] is prone to local minima, Method of Moments [1] requires separate analytic calculations for different variants of the model, cannot handle local spike history dependence. Both assume stationary linear Gaussian dynamics.

Nuclear Norm Heuristic

In the absence of spike history term, the matrix Y of pre-nonlinearity rates is *low-rank*. One way to find the subspace of dynamics might be to maximize the likelihood of the data with a penalty for the rank of Y

$$\min_Y \operatorname{rank}(Y) - \lambda \sum_i \log p(s_t|y_t)$$

But this is not convex! We can replace the rank condition with a convex function of Y that leads to low-rank solutions: the sum of singular values or *nuclear norm*

$$\min_Y ||Y||_* - \lambda \sum_t \log p(s_t|y_t)$$

This is a variant of *exponential family PCA* for point processes [2,6]. Unlike other approaches, this allows for a globally optimal solution.

Alternating Direction Method of Multipliers

How do we perform this minimization? Add auxiliary variable Z to separate smooth log likelihood and nuclear norm term. Add Lagrange multipliers U to enforce equality between them, and augment Lagrangian with the difference between Y and Z to make the optimization more robust.

$$\mathcal{L}(Y,Z,U) = ||Z||_* - \lambda \sum_t \log p(s_t|y_t) + \langle U, Y-Z \rangle + \frac{\rho}{2}||Y-Z||_2^2$$

$$Y_{k+1} = \arg\min_Y \mathcal{L}(Y, Z_k, U_k) \qquad \text{Newton's Method: } \mathcal{O}(NT)$$

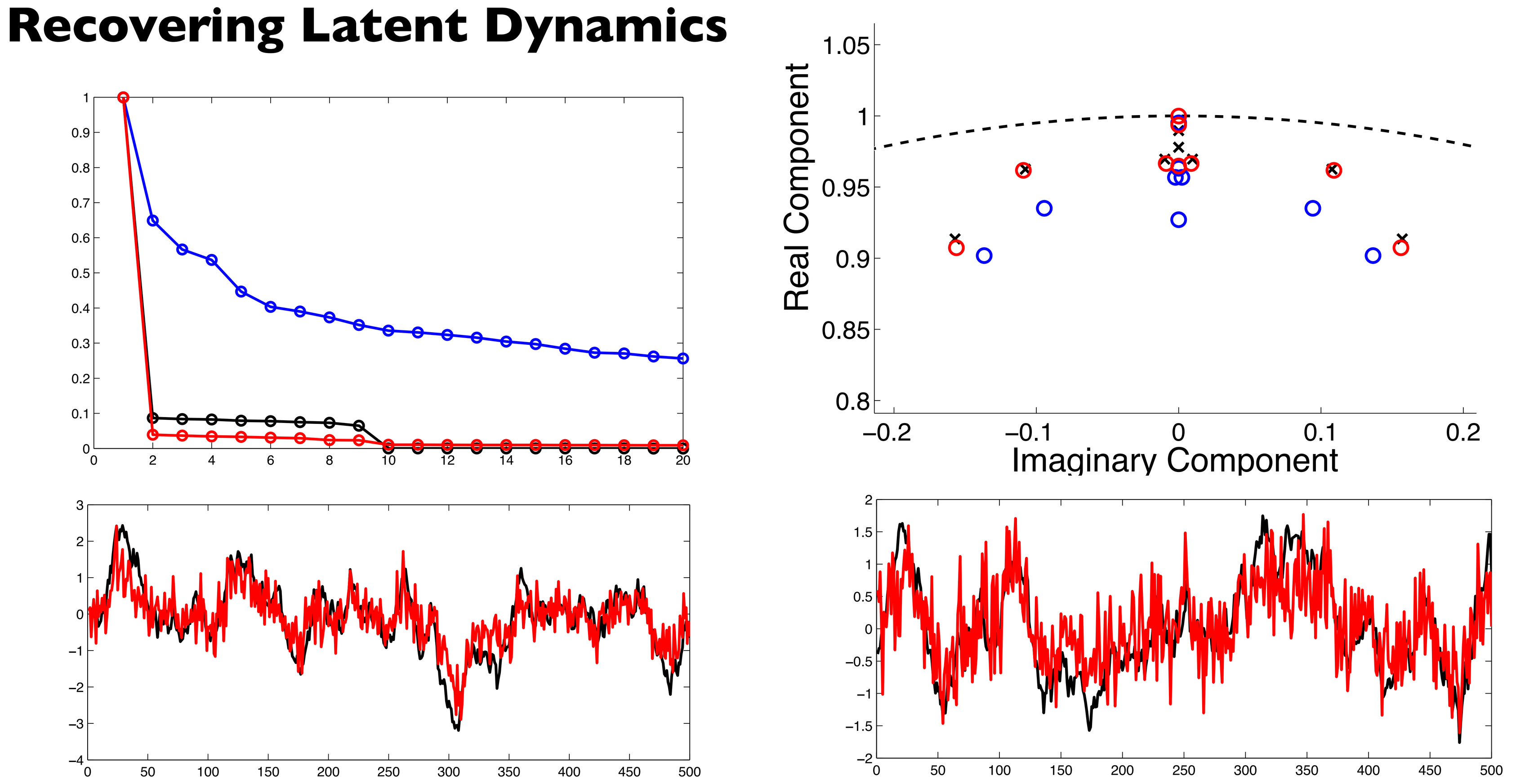
$$Z_{k+1} = \arg\min_Z \mathcal{L}(Y_{k+1}, Z, U_k) \qquad \text{SVD threshold: } \mathcal{O}(N^2T)$$

$$U_{k+1} = U_k + \rho(Y_{k+1} - Z_{k+1}) \qquad \text{Dual gradient ascent}$$

Guaranteed convergence to the global optimum! Applied to linear system identification with Gaussian noise in [4]. This is the first application we know of to other concave log likelihoods.

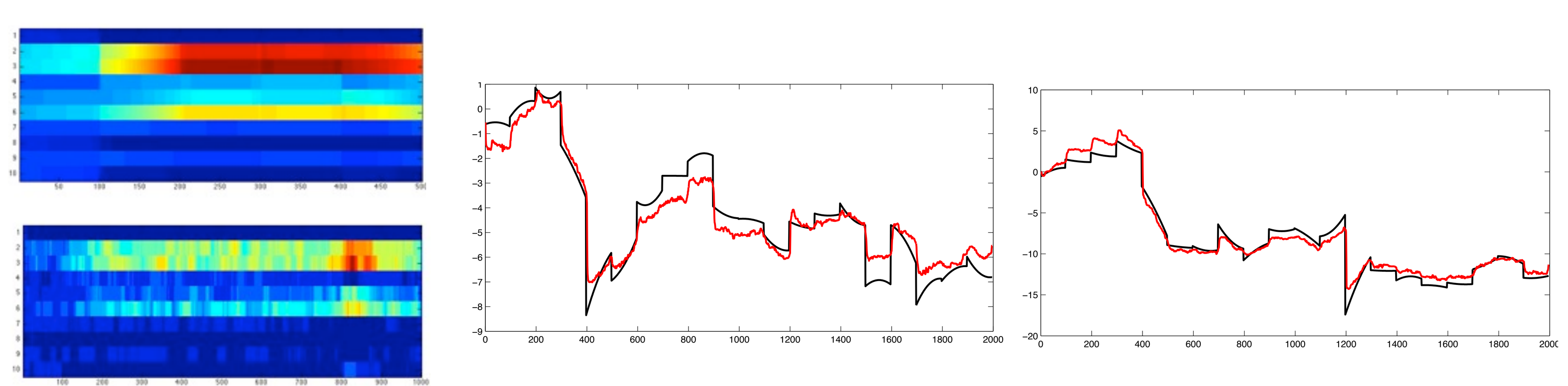
Experiments

Recovering Latent Dynamics



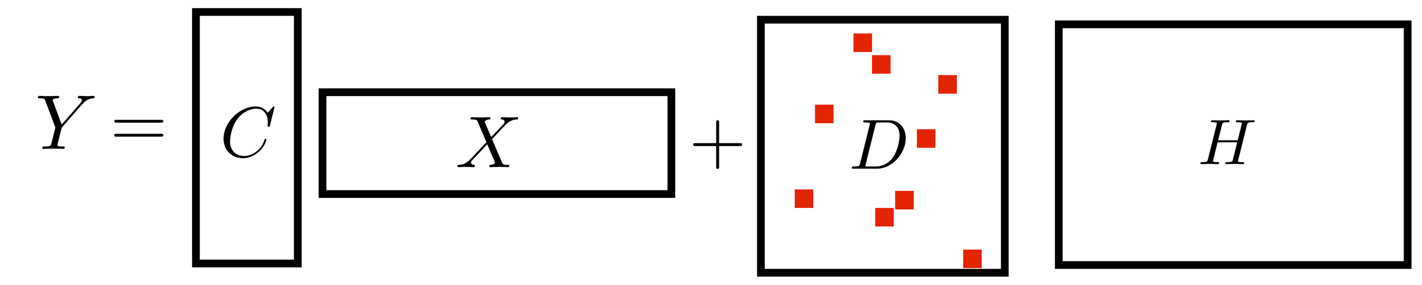
Synthetic data, linear-Gaussian system, 500 neurons, 8 dimensional dynamics, 10000 bins, f(x) = exp(x). Top left: scaled spectrum of the true matrix Y (black), spike matrix S (blue) and recovered matrix $\hat{Y}$ (red). Note the spectrum of $\hat{Y}$ more closely match the shape of Y. Top right: Eigenvalues of the true dynamics matrix A (black) and matrices recovered by NN minimization (red) and subspace identification (blue). Circle is the radius of stability. Note the biases in subspace ID due to the nonlinearity. Bottom: true (black) vs recovered (red) dynamics for 2 of the latent dimensions.

Non-Gaussian Non-stationary Dynamics



Synthetic data, 200 neurons, 2 latent dimensions, 2000 bins, soft nonlinearity. The dynamics are strongly non-Gaussian and non-stationary, but can still be recovered if the number of neurons is large enough relative to the rank of the dynamics. Left top: true firing rate for a subset of neurons. Left bottom: recovered firing rate plus median filtering. Right: true latent dynamics (black) vs recovered dynamics (red) after median filtering for two dimensions. In principle this is possible even if certain units drop out for some stretch.

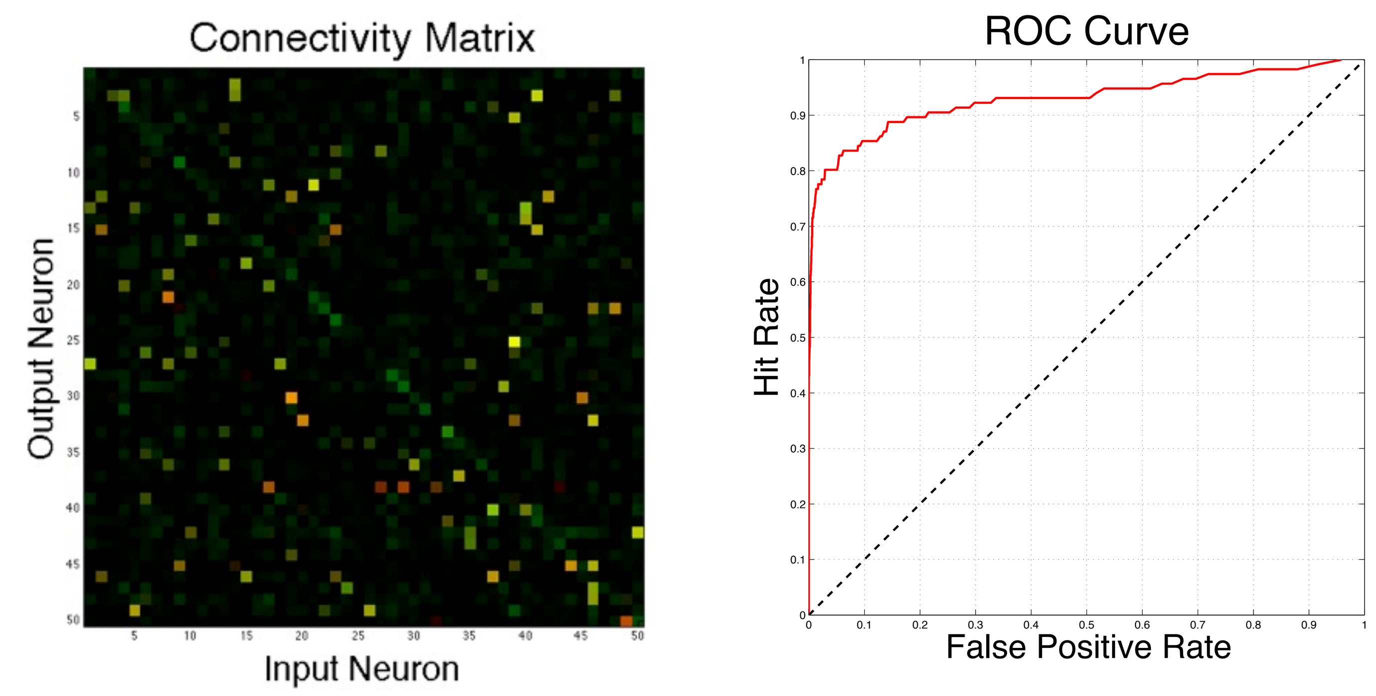
Separating Dynamics from Connections



If we include the linear dependence on spike history, Y is no longer low-rank. However, in large networks the history matrix is likely to be *sparse* and the connectivity can be separated from the low-rank dynamics by *stable principal component pursuit* (SPCP) [7]:

$$\min_{L,S} ||L||_* + \gamma||SH^\dagger||_1 - \lambda \sum_t p(s_t|y_t, s_{t-k:t-1})$$

such that $L + S = Y$ and $H^\dagger$ is the pseudoinverse of the spike history. Unlike existing methods to separate local connectivity from global dynamics based on EM [2,6], this is a convex problem.



SPCP performance on model data. 50 neurons, 10 latent dimensions, 5% connectivity, 5000 bins, max firing rate = 25/bin, soft nonlinearity.
Left: Comparison of true and recovered connectivity on model data. False positives in green, false negatives in red, hits in yellow. Intensity is synaptic strength.
Right: ROC curve showing 90% synapses recovered with only 20% false positives.

Conclusions

-Nuclear norm minimization provides a convex heuristic for problems with low-dimensional solutions - no local minima!
-Different noise models, nonlinearities can easily be swapped in.
-Numerical experiments show that as the number of neurons grows, accuracy increases, even for non-stationary dynamics.
-In certain regimes, sparse local connectivity can be separated from low-dimensional global influences.

References

[1] Buesing L. et al, *Advances in Neural Information Processing Systems*, 25, 2012.
[2] Collins M. et al, *Advances in Neural Information Processing Systems*, 14, 2001.
[3] Kulkarni J.E. and Paninski L., *Network: Computation in Neural Systems*, 18(4):375-307, 2007.
[4] Liu Z. et al, *Systems and Control Letters*, submitted.
[5] Macke J.H. et al, *Advances in Neural Information Processing Systems*, 24, 2011.
[6] Solo V. and Pasha S.A., *Neural Computation* 25(1):101-22, 2013
[7] Smith A.C. and Brown E.N., *Neural Computation*, 15(5):965-991, 2003.
[8] Vidne M. et al, *Journal of Computational Neuroscience*, 33(1):97-121, 2012.
[9] Zhou Z. et al, *IEEE International Symposium on Information Theory*, 2010