Unsupervised Discovery of Temporal Structure in Noisy Data with Dynamical Components Analysis

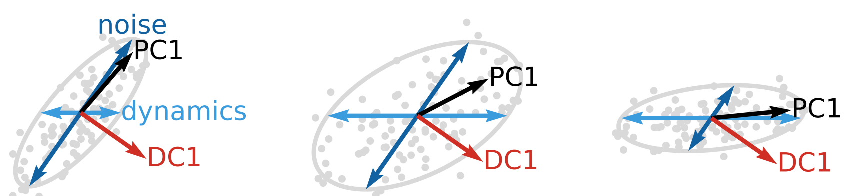
David G. Clark[1,2], Jesse A. Livezey[2,3], Kristofer E. Bouchard[2,3,4]

1. Center for Theoretical Neuroscience, Columbia University, 2. Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory,
3. Redwood Center for Theoretical Neuroscience, University of California, Berkeley, 4. Helen Wills Neuroscience Institute, University of California, Berkeley

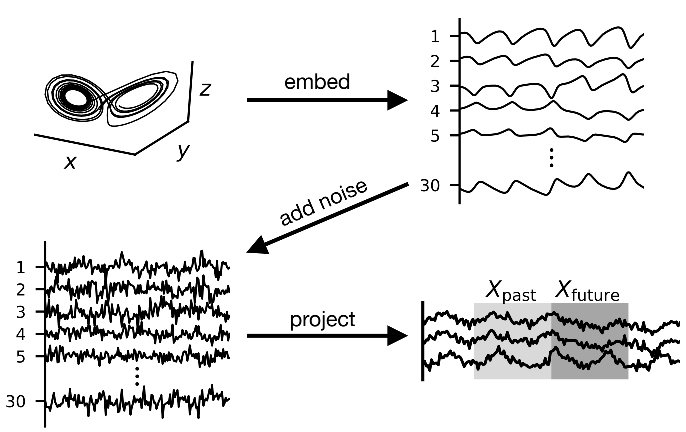
Overview

- High-dimensional time series datasets abound in many fields, e.g. neuroscience.
- Linear dimensionality reduction methods often aid the interpretation of such data.
- However, methods that do not take time into account, such as PCA, can miss dynamical structure.
- We tackle this problem by introducing Dynamical Components Analysis (DCA), a linear dimensionality reduction method with an information-theoretic objective function that specifically extracts dynamical structure.

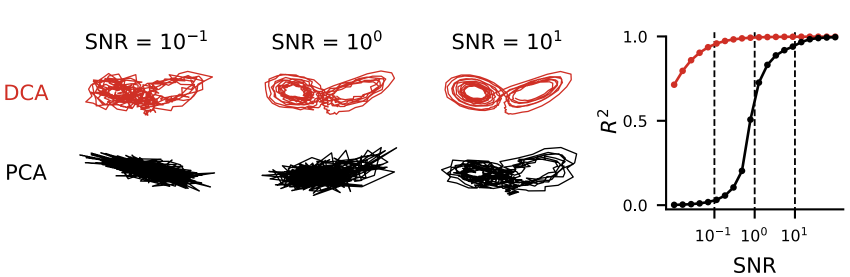
Dynamics and noise both contribute to variance



PCA, but not DCA, fails to extract an embedded Lorenz attractor with spatially structured noise



$$\text{SNR} = \frac{\text{variance of 1st dynamics PC}}{\text{variance of 1st noise PC}}$$



Predictive information (PI) measures dynamical content

$$\begin{aligned} I_T^{\text{pred}}(X) &= H\left(X_{\text{future}}\right) - H\left(X_{\text{future}} \mid X_{\text{past}}\right) \\ &= H\left(X_{\text{past}}\right) + H\left(X_{\text{future}}\right) - H\left(X_{\text{past}}, X_{\text{future}}\right) \\ &= 2H_X(T) - H_X(2T) \end{aligned}$$

Predictive information, Bialek and Tishby (1999)

- Mutual information is **difficult to estimate**
- Estimate must be **differentiable** in a projection matrix

Gaussian approximation enables PI optimization

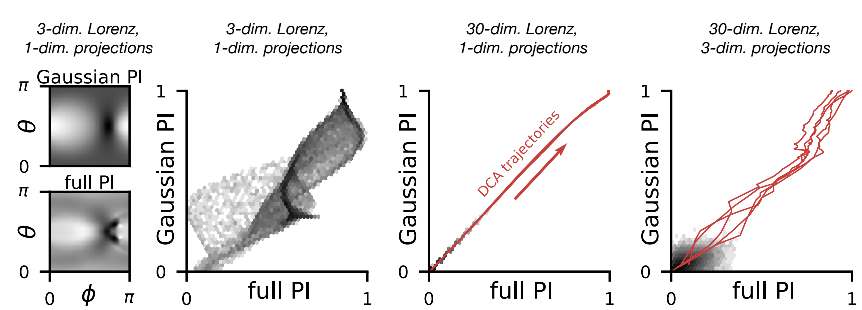
$$\begin{aligned} I_T^{\text{pred}}(X) &= 2H_X(T) - H_X(2T) \\ &= \log\left|\Sigma_T(X)\right| - \frac{1}{2}\log\left|\Sigma_{2T}(X)\right|. \end{aligned}$$

where

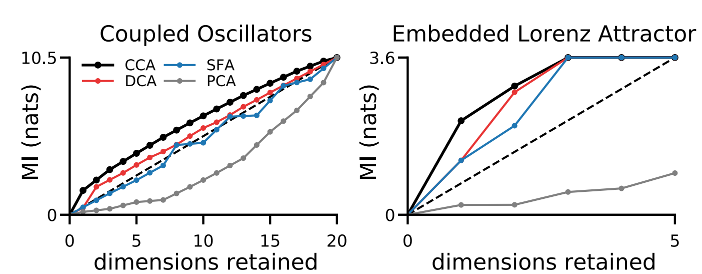
$$\Sigma_T(X) = \begin{pmatrix} C_0 & C_1 & \dots & C_{T-1} \\ C_1^T & C_0 & \dots & C_{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ C_{T-1}^T & C_{T-2}^T & \dots & C_0 \end{pmatrix} \qquad C_{\Delta t} = \left\langle x_t x_{t+\Delta t}^T \right\rangle_t$$

To compute $\Sigma_T(Y)$ for the projected time series Y given by $\mathbf{y}(t) = V^T \mathbf{x}(t)$, send $C_{\Delta t} \rightarrow V^T C_{\Delta t} V$

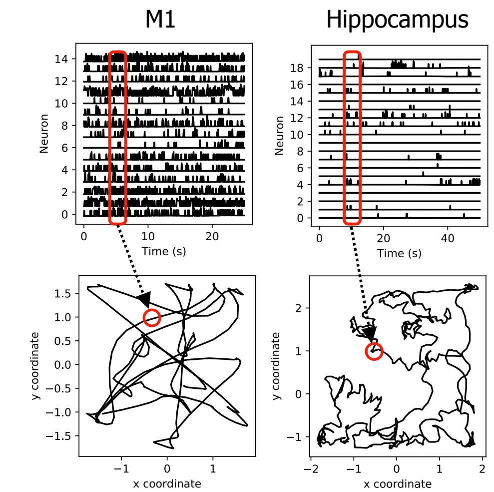
2nd-order statistics are sufficient for optimizing PI

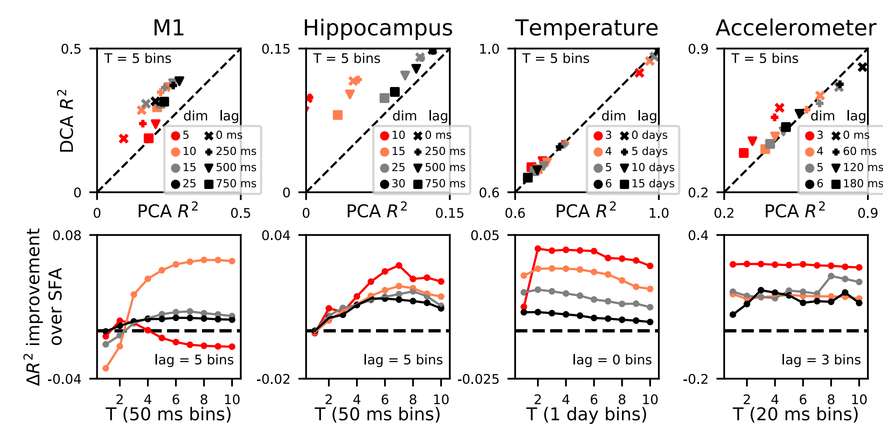


DCA captures more (Gaussian) past-future mutual information than PCA and Slow Feature Analysis

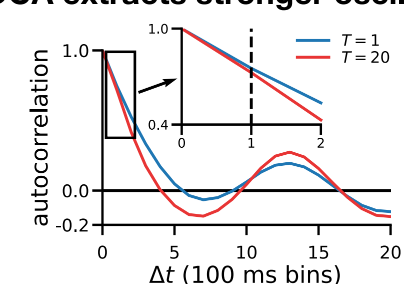


DCA subspaces outperform those found by PCA and Slow Feature Analysis at decoding and forecasting





On M1 data, DCA extracts stronger oscillations for $T > 1$



Future directions

- Extending to non-stationary time series
- Preprocessing step for computationally demanding methods (e.g., LFADS)
- Sparse feature selection via L1 regularization
- Kernel DCA via the kernel trick

This work was supported by the Laboratory Directed Research and Development Program of Lawrence Berkeley National Laboratory

SCAN ME