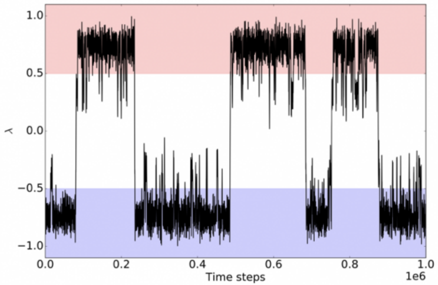
Learning “slow” manifolds from dynamical data

F. Noé[1]

Deep Learning Classes, FU Berlin 2018

Objective: find "collective coordinates" in which events are rare



Collective coordinate: Linear projection $u = \mathbf{x}_t^\top \mathbf{w}$ or nonlinear function $u = f(\mathbf{x}_t)$.

Nomenclature

- Time series $\{\mathbf{x}(1), \ldots, \mathbf{x}(T)\}$. Time-lagged data subsets:

$$\mathbf{X} = \begin{bmatrix} x_1(t=1) & \cdots & x_D(t=1) \\ \vdots & & \vdots \\ x_1(t=T-\tau) & \cdots & x_D(t=T-\tau) \end{bmatrix}$$

$$\mathbf{Y} = \begin{bmatrix} x_1(t=\tau) & \cdots & x_D(t=\tau) \\ \vdots & & \vdots \\ x_1(t=T) & \cdots & x_D(t=T) \end{bmatrix}$$

- Assume that the data is mean-free:

$$\sum_{t=1}^{T-\tau} \mathbf{x}(t) = \mathbf{0} \qquad \sum_{t=1}^{T-\tau} \mathbf{y}(t) = \mathbf{0}$$

Otherwise remove the mean in $\mathbf{X}$ and $\mathbf{Y}$ *separately*

Nomenclature

- Time series $\{\mathbf{x}(1), ..., \mathbf{x}(T)\}$. Time-lagged data subsets:

$$\begin{aligned}
\mathbf{X} &= \begin{bmatrix} x_1(t=1) & \cdots & x_D(t=1) \\ \vdots & & \vdots \\ x_1(t=T-\tau) & \cdots & x_D(t=T-\tau) \end{bmatrix} \\
\mathbf{Y} &= \begin{bmatrix} x_1(t=\tau) & \cdots & x_D(t=\tau) \\ \vdots & & \vdots \\ x_1(t=T) & \cdots & x_D(t=T) \end{bmatrix}
\end{aligned}$$

- Assume that the data is mean-free:

$$\sum_{t=1}^{T-\tau} \mathbf{x}(t) = \mathbf{0} \qquad \sum_{t=1}^{T-\tau} \mathbf{y}(t) = \mathbf{0}$$

Otherwise remove the mean in **X** and **Y** *separately*

Whitening

- Compute covariance matrix and perform PCA:

$$\mathbf{C} = \mathbf{X}^\top \mathbf{X} \qquad \mathbf{CU} = \mathbf{U\Sigma}^2$$

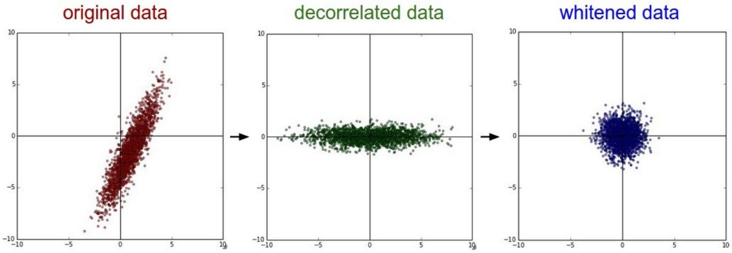
- Decorrelate data:

$$\mathbf{X}' = \mathbf{XU}$$

- Whiten data:

$$\tilde{\mathbf{X}} = \mathbf{XU\Sigma}^{-1} = \mathbf{XC}^{-\frac{1}{2}}$$

which uses: $\mathbf{C}^{-1} = \mathbf{U\Sigma}^{-1}\mathbf{\Sigma}^{-1}\mathbf{U}^\top = \mathbf{C}^{-\frac{1}{2}}\mathbf{C}^{-\frac{1}{2}}$



[1]

[1] From AI-Odyssey.com

Whitening

- Compute covariance matrix and perform PCA:

$$\mathbf{C} = \mathbf{X}^\top \mathbf{X} \qquad \mathbf{C}\mathbf{U} = \mathbf{U}\boldsymbol{\Sigma}^2$$

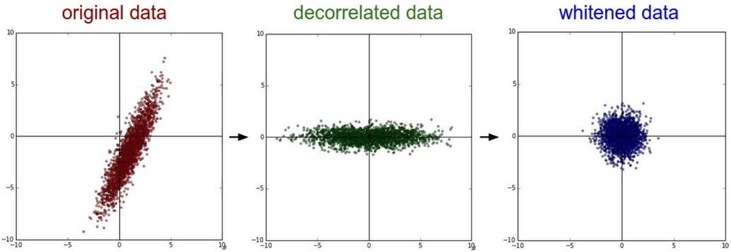
- Decorrelate data:

$$\mathbf{X}' = \mathbf{X}\mathbf{U}$$

- Whiten data:

$$\tilde{\mathbf{X}} = \mathbf{X}\mathbf{U}\boldsymbol{\Sigma}^{-1} = \mathbf{X}\mathbf{C}^{-\frac{1}{2}}$$

which uses: $\mathbf{C}^{-1} = \mathbf{U}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^{-1}\mathbf{U}^\top = \mathbf{C}^{-\frac{1}{2}}\mathbf{C}^{-\frac{1}{2}}$



[1]

[1] From AI-Odyssey.com

Whitening

- Compute covariance matrix and perform PCA:

$$\mathbf{C} = \mathbf{X}^\top \mathbf{X} \qquad \mathbf{C}\mathbf{U} = \mathbf{U}\boldsymbol{\Sigma}^2$$

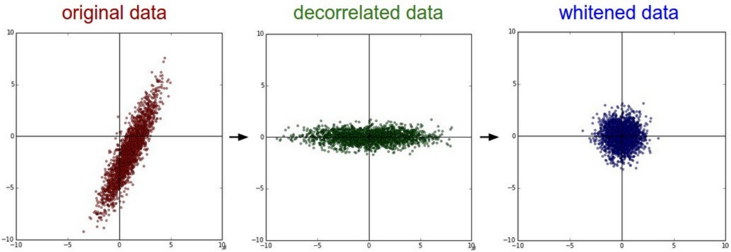
- Decorrelate data:

$$\mathbf{X}' = \mathbf{X}\mathbf{U}$$

- Whiten data:

$$\tilde{\mathbf{X}} = \mathbf{X}\mathbf{U}\boldsymbol{\Sigma}^{-1} = \mathbf{X}\mathbf{C}^{-\frac{1}{2}}$$

which uses: $\mathbf{C}^{-1} = \mathbf{U}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^{-1}\mathbf{U}^\top = \mathbf{C}^{-\frac{1}{2}}\mathbf{C}^{-\frac{1}{2}}$



[1]

[1] From AI-Odyssey.com

Summary – Data Preparation

- Assume coordinates are mean free (otherwise remove mean):

$$\sum_{t=1}^{T-\tau} \mathbf{x}_t = 0 \qquad \sum_{t=1}^{T-\tau} \mathbf{y}_t = 0$$

- Covariance matrices:

$$\mathbf{C}_{xx} = \frac{1}{T-\tau} \sum_{t=1}^{T-\tau} \mathbf{x}_t \mathbf{x}_t^\top = \frac{1}{T-\tau} \mathbf{X}^\top \mathbf{X}$$

$$\mathbf{C}_{xy} = \frac{1}{T-\tau} \sum_{t=1}^{T-\tau} \mathbf{x}_t \mathbf{y}_t^\top = \frac{1}{T-\tau} \mathbf{X}^\top \mathbf{Y}$$

$$\mathbf{C}_{yy} = \frac{1}{T-\tau} \sum_{t=1}^{T-\tau} \mathbf{y}_t \mathbf{y}_t^\top = \frac{1}{T-\tau} \mathbf{Y}^\top \mathbf{Y}$$

- Whiten data:

$$\tilde{\mathbf{x}}_t = \mathbf{C}_{xx}^{-\frac{1}{2}} \mathbf{x}_t$$

$$\tilde{\mathbf{y}}_t = \mathbf{C}_{yy}^{-\frac{1}{2}} \mathbf{y}_t$$

Summary – Data Preparation

- Assume coordinates are mean free (otherwise remove mean):

$$\sum_{t=1}^{T-\tau} \mathbf{x}_t = 0 \qquad \sum_{t=1}^{T-\tau} \mathbf{y}_t = 0$$

- Covariance matrices:

$$\mathbf{C}_{xx} = \frac{1}{T-\tau} \sum_{t=1}^{T-\tau} \mathbf{x}_t \mathbf{x}_t^\top = \frac{1}{T-\tau} \mathbf{X}^\top \mathbf{X}$$

$$\mathbf{C}_{xy} = \frac{1}{T-\tau} \sum_{t=1}^{T-\tau} \mathbf{x}_t \mathbf{y}_t^\top = \frac{1}{T-\tau} \mathbf{X}^\top \mathbf{Y}$$

$$\mathbf{C}_{yy} = \frac{1}{T-\tau} \sum_{t=1}^{T-\tau} \mathbf{y}_t \mathbf{y}_t^\top = \frac{1}{T-\tau} \mathbf{Y}^\top \mathbf{Y}$$

- Whiten data:

$$\tilde{\mathbf{x}}_t = \mathbf{C}_{xx}^{-\frac{1}{2}} \mathbf{x}_t$$

$$\tilde{\mathbf{y}}_t = \mathbf{C}_{yy}^{-\frac{1}{2}} \mathbf{y}_t$$

Summary – Data Preparation

- Assume coordinates are mean free (otherwise remove mean):

$$\sum_{t=1}^{T-\tau} \mathbf{x}_t = 0 \qquad \sum_{t=1}^{T-\tau} \mathbf{y}_t = 0$$

- Covariance matrices:

$$\mathbf{C}_{xx} = \frac{1}{T-\tau} \sum_{t=1}^{T-\tau} \mathbf{x}_t \mathbf{x}_t^\top = \frac{1}{T-\tau} \mathbf{X}^\top \mathbf{X}$$

$$\mathbf{C}_{xy} = \frac{1}{T-\tau} \sum_{t=1}^{T-\tau} \mathbf{x}_t \mathbf{y}_t^\top = \frac{1}{T-\tau} \mathbf{X}^\top \mathbf{Y}$$

$$\mathbf{C}_{yy} = \frac{1}{T-\tau} \sum_{t=1}^{T-\tau} \mathbf{y}_t \mathbf{y}_t^\top = \frac{1}{T-\tau} \mathbf{Y}^\top \mathbf{Y}$$

- Whiten data:

$$\tilde{\mathbf{x}}_t = \mathbf{C}_{xx}^{-\frac{1}{2}} \mathbf{x}_t$$

$$\tilde{\mathbf{y}}_t = \mathbf{C}_{yy}^{-\frac{1}{2}} \mathbf{y}_t$$

“Best” Projection of Dynamical Data

- Seek **Encoding** (projection) and **Decoding**:

$$\mathbf{z}_t = \tilde{\mathbf{E}}\tilde{\mathbf{x}}_t$$
$$\tilde{\mathbf{y}}_t^* = \tilde{\mathbf{D}}\mathbf{z}_t = \tilde{\mathbf{K}}_d\tilde{\mathbf{x}}_t$$

with $\tilde{\mathbf{E}} = \mathbb{R}^{m\times n}$, $\tilde{\mathbf{D}} = \mathbb{R}^{n\times m}$ and the rank-d transition matrix:

$$\tilde{\mathbf{K}}_d = \tilde{\mathbf{D}}\tilde{\mathbf{E}}$$

- **Aim**: As for PCA, we can formulate two different optimization strategies: (i) minimizing a suitable error, and (ii) maximizing a suitable variance.
- Here: **minimize prediction error** by solving regression problem:

$$\min_{\tilde{\mathbf{K}}_d} \left\|\tilde{\mathbf{Y}} - \tilde{\mathbf{Y}}^*\right\|_F^2 = \min_{\tilde{\mathbf{K}}_d} \left\|\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\tilde{\mathbf{K}}_d^\top\right\|_F^2$$

- Seek **Encoding** (projection) and **Decoding**:

$$\mathbf{z}_t = \tilde{\mathbf{E}}\tilde{\mathbf{x}}_t$$
$$\tilde{\mathbf{y}}_t^* = \tilde{\mathbf{D}}\mathbf{z}_t = \tilde{\mathbf{K}}_d\tilde{\mathbf{x}}_t$$

with $\tilde{\mathbf{E}} = \mathbb{R}^{m\times n}$, $\tilde{\mathbf{D}} = \mathbb{R}^{n\times m}$ and the rank-d transition matrix:

$$\tilde{\mathbf{K}}_d = \tilde{\mathbf{D}}\tilde{\mathbf{E}}$$

- **Aim**: As for PCA, we can formulate two different optimization strategies: (i) minimizing a suitable error, and (ii) maximizing a suitable variance.
- Here: **minimize prediction error** by solving regression problem:

$$\min_{\tilde{\mathbf{K}}_d} \left\| \tilde{\mathbf{Y}} - \tilde{\mathbf{Y}}^* \right\|_F^2 = \min_{\tilde{\mathbf{K}}_d} \left\| \tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\tilde{\mathbf{K}}_d^\top \right\|_F^2$$

“Best” Projection of Dynamical Data

- Seek **Encoding** (projection) and **Decoding**:

$$\begin{aligned} \mathbf{z}_t &= \tilde{\mathbf{E}}\tilde{\mathbf{x}}_t \\ \tilde{\mathbf{y}}_t^* &= \tilde{\mathbf{D}}\mathbf{z}_t = \tilde{\mathbf{K}}_d\tilde{\mathbf{x}}_t \end{aligned}$$

with $\tilde{\mathbf{E}} = \mathbb{R}^{m\times n}$, $\tilde{\mathbf{D}} = \mathbb{R}^{n\times m}$ and the rank-d transition matrix:

$$\tilde{\mathbf{K}}_d = \tilde{\mathbf{D}}\tilde{\mathbf{E}}$$

- **Aim**: As for PCA, we can formulate two different optimization strategies: (i) minimizing a suitable error, and (ii) maximizing a suitable variance.
- Here: **minimize prediction error** by solving regression problem:

$$\min_{\tilde{\mathbf{K}}_d} \left\|\tilde{\mathbf{Y}} - \tilde{\mathbf{Y}}^*\right\|_F^2 = \min_{\tilde{\mathbf{K}}_d} \left\|\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\tilde{\mathbf{K}}_d^\top\right\|_F^2$$

- Full-rank solution (using $\tilde{\mathbf{X}}^\top\tilde{\mathbf{X}} = (T-\tau)\mathbf{I}$):

$$\tilde{\mathbf{K}}^\top = \left(\tilde{\mathbf{X}}^\top\tilde{\mathbf{X}}\right)^{-1}\tilde{\mathbf{X}}^\top\tilde{\mathbf{Y}} = \frac{1}{T-\tau}\tilde{\mathbf{X}}^\top\tilde{\mathbf{Y}}$$

- Using definition of covariance matrices:

$$\tilde{\mathbf{K}}^\top = \mathbf{C}_{xx}^{-\frac{1}{2}}\mathbf{C}_{xy}\mathbf{C}_{yy}^{-\frac{1}{2}}$$

- Best low-rank approximation:

$$\tilde{\mathbf{K}}_d^\top = \mathrm{SVD}\left(\mathbf{C}_{xx}^{-\frac{1}{2}}\mathbf{C}_{xy}\mathbf{C}_{yy}^{-\frac{1}{2}}\right) = \tilde{\mathbf{U}}_d\mathbf{\Sigma}_d\tilde{\mathbf{V}}_d^\top$$

- Optimal encoder and decoder:

$$\tilde{\mathbf{E}}^\top = \tilde{\mathbf{U}}_d \quad \tilde{\mathbf{D}}^\top = \mathbf{\Sigma}_d\tilde{\mathbf{V}}_d^\top$$

  and defining $\mathbf{U}_d = \mathbf{C}_{xx}^{-\frac{1}{2}}\tilde{\mathbf{U}}_d$, $\mathbf{V}_d = \mathbf{C}_{yy}^{-\frac{1}{2}}\tilde{\mathbf{V}}_d$ the non-whitened versions:

$$\mathbf{E}^\top = \mathbf{U}_d \quad \mathbf{D}^\top = \mathbf{\Sigma}_d\mathbf{V}_d^\top$$

- **Time-lagged canonical covariance analysis (TCCA)**: H. Wu and F. Noé: Variational approach for learning Markov processes from time series data. arXiv:1707.04659 (2017).

"Best" Projection of Dynamical Data: TCCA

- Full-rank solution (using $\tilde{\mathbf{X}}^\top\tilde{\mathbf{X}} = (T-\tau)\mathbf{I}$):

$$\tilde{\mathbf{K}}^\top = \left(\tilde{\mathbf{X}}^\top\tilde{\mathbf{X}}\right)^{-1}\tilde{\mathbf{X}}^\top\tilde{\mathbf{Y}} = \frac{1}{T-\tau}\tilde{\mathbf{X}}^\top\tilde{\mathbf{Y}}$$

- Using definition of covariance matrices:

$$\tilde{\mathbf{K}}^\top = \mathbf{C}_{xx}^{-\frac{1}{2}}\mathbf{C}_{xy}\mathbf{C}_{yy}^{-\frac{1}{2}}$$

- Best low-rank approximation:

$$\tilde{\mathbf{K}}_d^\top = \mathrm{SVD}\left(\mathbf{C}_{xx}^{-\frac{1}{2}}\mathbf{C}_{xy}\mathbf{C}_{yy}^{-\frac{1}{2}}\right) = \tilde{\mathbf{U}}_d\mathbf{\Sigma}_d\tilde{\mathbf{V}}_d^\top$$

- Optimal encoder and decoder:

$$\tilde{\mathbf{E}}^\top = \tilde{\mathbf{U}}_d \quad \tilde{\mathbf{D}}^\top = \mathbf{\Sigma}_d\tilde{\mathbf{V}}_d^\top$$

  and defining $\mathbf{U}_d = \mathbf{C}_{xx}^{-\frac{1}{2}}\tilde{\mathbf{U}}_d$, $\mathbf{V}_d = \mathbf{C}_{yy}^{-\frac{1}{2}}\tilde{\mathbf{V}}_d$ the non-whitened versions:

$$\mathbf{E}^\top = \mathbf{U}_d \quad \mathbf{D}^\top = \mathbf{\Sigma}_d\mathbf{V}_d^\top$$

- **Time-lagged canonical covariance analysis (TCCA)**: H. Wu and F. Noé: Variational approach for learning Markov processes from time series data. arXiv:1707.04659 (2017).

"Best" Projection of Dynamical Data: TCCA

- Full-rank solution (using $\tilde{\mathbf{X}}^\top \tilde{\mathbf{X}} = (T-\tau)\mathbf{I}$):

$$\tilde{\mathbf{K}}^\top = \left(\tilde{\mathbf{X}}^\top \tilde{\mathbf{X}}\right)^{-1} \tilde{\mathbf{X}}^\top \tilde{\mathbf{Y}} = \frac{1}{T-\tau} \tilde{\mathbf{X}}^\top \tilde{\mathbf{Y}}$$

- Using definition of covariance matrices:

$$\tilde{\mathbf{K}}^\top = \mathbf{C}_{xx}^{-\frac{1}{2}} \mathbf{C}_{xy} \mathbf{C}_{yy}^{-\frac{1}{2}}$$

- Best low-rank approximation:

$$\tilde{\mathbf{K}}_d^\top = \mathrm{SVD}\left(\mathbf{C}_{xx}^{-\frac{1}{2}} \mathbf{C}_{xy} \mathbf{C}_{yy}^{-\frac{1}{2}}\right) = \tilde{\mathbf{U}}_d \mathbf{\Sigma}_d \tilde{\mathbf{V}}_d^\top$$

- Optimal encoder and decoder:

$$\tilde{\mathbf{E}}^\top = \tilde{\mathbf{U}}_d \quad \tilde{\mathbf{D}}^\top = \mathbf{\Sigma}_d \tilde{\mathbf{V}}_d^\top$$

and defining $\mathbf{U}_d = \mathbf{C}_{xx}^{-\frac{1}{2}} \tilde{\mathbf{U}}_d$, $\mathbf{V}_d = \mathbf{C}_{yy}^{-\frac{1}{2}} \tilde{\mathbf{V}}_d$ the non-whitened versions:

$$\mathbf{E}^\top = \mathbf{U}_d \quad \mathbf{D}^\top = \mathbf{\Sigma}_d \mathbf{V}_d^\top$$

- **Time-lagged canonical covariance analysis (TCCA)**: H. Wu and F. Noé: Variational approach for learning Markov processes from time series data. arXiv:1707.04659 (2017).

- Full-rank solution (using $\tilde{\mathbf{X}}^\top\tilde{\mathbf{X}} = (T-\tau)\mathbf{I}$):

$$\tilde{\mathbf{K}}^\top = \left(\tilde{\mathbf{X}}^\top\tilde{\mathbf{X}}\right)^{-1}\tilde{\mathbf{X}}^\top\tilde{\mathbf{Y}} = \frac{1}{T-\tau}\tilde{\mathbf{X}}^\top\tilde{\mathbf{Y}}$$

- Using definition of covariance matrices:

$$\tilde{\mathbf{K}}^\top = \mathbf{C}_{xx}^{-\frac{1}{2}}\mathbf{C}_{xy}\mathbf{C}_{yy}^{-\frac{1}{2}}$$

- Best low-rank approximation:

$$\tilde{\mathbf{K}}_d^\top = \mathrm{SVD}\left(\mathbf{C}_{xx}^{-\frac{1}{2}}\mathbf{C}_{xy}\mathbf{C}_{yy}^{-\frac{1}{2}}\right) = \tilde{\mathbf{U}}_d\mathbf{\Sigma}_d\tilde{\mathbf{V}}_d^\top$$

- Optimal encoder and decoder:

$$\tilde{\mathbf{E}}^\top = \tilde{\mathbf{U}}_d \quad \tilde{\mathbf{D}}^\top = \mathbf{\Sigma}_d\tilde{\mathbf{V}}_d^\top$$

  and defining $\mathbf{U}_d = \mathbf{C}_{xx}^{-\frac{1}{2}}\tilde{\mathbf{U}}_d$, $\mathbf{V}_d = \mathbf{C}_{yy}^{-\frac{1}{2}}\tilde{\mathbf{V}}_d$ the non-whitened versions:

$$\mathbf{E}^\top = \mathbf{U}_d \quad \mathbf{D}^\top = \mathbf{\Sigma}_d\mathbf{V}_d^\top$$

- **Time-lagged canonical covariance analysis (TCCA)**: H. Wu and F. Noé: Variational approach for learning Markov processes from time series data. arXiv:1707.04659 (2017).

"Best" Projection of Dynamical Data: TCCA

- Full-rank solution (using $\tilde{\mathbf{X}}^\top \tilde{\mathbf{X}} = (T - \tau)\mathbf{I}$):

$$\tilde{\mathbf{K}}^\top = \left(\tilde{\mathbf{X}}^\top \tilde{\mathbf{X}}\right)^{-1} \tilde{\mathbf{X}}^\top \tilde{\mathbf{Y}} = \frac{1}{T - \tau} \tilde{\mathbf{X}}^\top \tilde{\mathbf{Y}}$$

- Using definition of covariance matrices:

$$\tilde{\mathbf{K}}^\top = \mathbf{C}_{xx}^{-\frac{1}{2}} \mathbf{C}_{xy} \mathbf{C}_{yy}^{-\frac{1}{2}}$$

- Best low-rank approximation:

$$\tilde{\mathbf{K}}_d^\top = \mathrm{SVD}\left(\mathbf{C}_{xx}^{-\frac{1}{2}} \mathbf{C}_{xy} \mathbf{C}_{yy}^{-\frac{1}{2}}\right) = \tilde{\mathbf{U}}_d \mathbf{\Sigma}_d \tilde{\mathbf{V}}_d^\top$$

- Optimal encoder and decoder:

$$\tilde{\mathbf{E}}^\top = \tilde{\mathbf{U}}_d \quad \tilde{\mathbf{D}}^\top = \mathbf{\Sigma}_d \tilde{\mathbf{V}}_d^\top$$

  and defining $\mathbf{U}_d = \mathbf{C}_{xx}^{-\frac{1}{2}} \tilde{\mathbf{U}}_d$, $\mathbf{V}_d = \mathbf{C}_{yy}^{-\frac{1}{2}} \tilde{\mathbf{V}}_d$ the non-whitened versions:

$$\mathbf{E}^\top = \mathbf{U}_d \quad \mathbf{D}^\top = \mathbf{\Sigma}_d \mathbf{V}_d^\top$$

- **Time-lagged canonical covariance analysis (TCCA)**: H. Wu and F. Noé: Variational approach for learning Markov processes from time series data. arXiv:1707.04659 (2017).

Special case: reversible dynamics: TICA

- Consider both "forward" time $\{\mathbf{x}(1), \ldots, \mathbf{x}(T)\}$ and "backward" time $\{\mathbf{x}(T), \ldots, \mathbf{x}(1)\}$. Thus:

$$\mathbf{C}_{xx} = \mathbf{C}_{yy} = \mathbf{X}^\top \mathbf{X} + \mathbf{Y}^\top \mathbf{Y}^\top$$
$$\mathbf{C}_{xy} = \mathbf{C}_{yx} = \mathbf{X}^\top \mathbf{Y} + \mathbf{Y}^\top \mathbf{X}^\top$$

- Result:

$$\mathrm{SVD}\left(\mathbf{C}_{xx}^{-\frac{1}{2}} \mathbf{C}_{xy} \mathbf{C}_{xx}^{-\frac{1}{2}}\right) = \mathrm{EVD}\left(\mathbf{C}_{xx}^{-\frac{1}{2}} \mathbf{C}_{xy} \mathbf{C}_{xx}^{-\frac{1}{2}}\right) = \tilde{\mathbf{U}}_d \mathbf{\Sigma}_d \tilde{\mathbf{U}}_d^\top$$

- We get the optimal encoder and decoder:

$$\tilde{\mathbf{E}}^\top = \tilde{\mathbf{U}}_d \quad \tilde{\mathbf{D}}^\top = \mathbf{\Sigma}_d \tilde{\mathbf{U}}_d^\top$$

and defining $\mathbf{U}_d = \mathbf{C}_{xx}^{-\frac{1}{2}} \tilde{\mathbf{U}}_d$ the non-whitened versions:

$$\mathbf{E}^\top = \mathbf{U}_d \quad \mathbf{D}^\top = \mathbf{\Sigma}_d \mathbf{U}_d^\top$$

- **Blind source separation**: Molgedey and Schuster. **Phys. Rev. Lett.** 72, 3634 (1994).
- **Time-lagged independent component analysis (TICA)**: G. Pérez-Hernández, F. Paul, T. Giorgino, G. De Fabritiis, F. Noé: Identification of slow molecular order parameters for Markov model construction. J. Chem. Phys. 139, 015102 (2013).

Special case: reversible dynamics: TICA

- Consider both "forward" time $\{\mathbf{x}(1), \ldots, \mathbf{x}(T)\}$ and "backward" time $\{\mathbf{x}(T), \ldots, \mathbf{x}(1)\}$. Thus:

$$\mathbf{C}_{xx} = \mathbf{C}_{yy} = \mathbf{X}^\top \mathbf{X} + \mathbf{Y}^\top \mathbf{Y}^\top$$
$$\mathbf{C}_{xy} = \mathbf{C}_{yx} = \mathbf{X}^\top \mathbf{Y} + \mathbf{Y}^\top \mathbf{X}^\top$$

- Result:

$$\mathrm{SVD}\left(\mathbf{C}_{xx}^{-\frac{1}{2}} \mathbf{C}_{xy} \mathbf{C}_{xx}^{-\frac{1}{2}}\right) = \mathrm{EVD}\left(\mathbf{C}_{xx}^{-\frac{1}{2}} \mathbf{C}_{xy} \mathbf{C}_{xx}^{-\frac{1}{2}}\right) = \tilde{\mathbf{U}}_d \boldsymbol{\Sigma}_d \tilde{\mathbf{U}}_d^\top$$

- We get the optimal encoder and decoder:

$$\tilde{\mathbf{E}}^\top = \tilde{\mathbf{U}}_d \quad \tilde{\mathbf{D}}^\top = \boldsymbol{\Sigma}_d \tilde{\mathbf{U}}_d^\top$$

and defining $\mathbf{U}_d = \mathbf{C}_{xx}^{-\frac{1}{2}} \tilde{\mathbf{U}}_d$ the non-whitened versions:

$$\mathbf{E}^\top = \mathbf{U}_d \quad \mathbf{D}^\top = \boldsymbol{\Sigma}_d \mathbf{U}_d^\top$$

- **Blind source separation**: Molgedey and Schuster. **Phys. Rev. Lett.** 72, 3634 (1994).
- **Time-lagged independent component analysis (TICA)**: G. Pérez-Hernández, F. Paul, T. Giorgino, G. De Fabritiis, F. Noé: Identification of slow molecular order parameters for Markov model construction. J. Chem. Phys. 139, 015102 (2013).

Special case: reversible dynamics: TICA

- Consider both "forward" time $\{\mathbf{x}(1),...,\mathbf{x}(T)\}$ and "backward" time $\{\mathbf{x}(T),...,\mathbf{x}(1)\}$. Thus:

$$\mathbf{C}_{xx} = \mathbf{C}_{yy} = \mathbf{X}^\top\mathbf{X} + \mathbf{Y}^\top\mathbf{Y}^\top$$
$$\mathbf{C}_{xy} = \mathbf{C}_{yx} = \mathbf{X}^\top\mathbf{Y} + \mathbf{Y}^\top\mathbf{X}^\top$$

- Result:

$$\mathrm{SVD}\left(\mathbf{C}_{xx}^{-\frac{1}{2}}\mathbf{C}_{xy}\mathbf{C}_{xx}^{-\frac{1}{2}}\right) = \mathrm{EVD}\left(\mathbf{C}_{xx}^{-\frac{1}{2}}\mathbf{C}_{xy}\mathbf{C}_{xx}^{-\frac{1}{2}}\right) = \tilde{\mathbf{U}}_d\boldsymbol{\Sigma}_d\tilde{\mathbf{U}}_d^\top$$

- We get the optimal encoder and decoder:

$$\tilde{\mathbf{E}}^\top = \tilde{\mathbf{U}}_d \quad \tilde{\mathbf{D}}^\top = \boldsymbol{\Sigma}_d\tilde{\mathbf{U}}_d^\top$$

and defining $\mathbf{U}_d = \mathbf{C}_{xx}^{-\frac{1}{2}}\tilde{\mathbf{U}}_d$ the non-whitened versions:

$$\mathbf{E}^\top = \mathbf{U}_d \quad \mathbf{D}^\top = \boldsymbol{\Sigma}_d\mathbf{U}_d^\top$$

- **Blind source separation**: Molgedey and Schuster. **Phys. Rev. Lett.** 72, 3634 (1994).
- **Time-lagged independent component analysis (TICA)**: G. Pérez-Hernández, F. Paul, T. Giorgino, G. De Fabritiis, F. Noé: Identification of slow molecular order parameters for Markov model construction. J. Chem. Phys. 139, 015102 (2013).

Special case: reversible dynamics: TICA

- Consider both "forward" time $\{\mathbf{x}(1), ..., \mathbf{x}(T)\}$ and "backward" time $\{\mathbf{x}(T), ..., \mathbf{x}(1)\}$. Thus:

$$\mathbf{C}_{xx} = \mathbf{C}_{yy} = \mathbf{X}^\top \mathbf{X} + \mathbf{Y}^\top \mathbf{Y}^\top$$
$$\mathbf{C}_{xy} = \mathbf{C}_{yx} = \mathbf{X}^\top \mathbf{Y} + \mathbf{Y}^\top \mathbf{X}^\top$$

- Result:

$$\mathrm{SVD}\left(\mathbf{C}_{xx}^{-\frac{1}{2}} \mathbf{C}_{xy} \mathbf{C}_{xx}^{-\frac{1}{2}}\right) = \mathrm{EVD}\left(\mathbf{C}_{xx}^{-\frac{1}{2}} \mathbf{C}_{xy} \mathbf{C}_{xx}^{-\frac{1}{2}}\right) = \tilde{\mathbf{U}}_d \mathbf{\Sigma}_d \tilde{\mathbf{U}}_d^\top$$

- We get the optimal encoder and decoder:

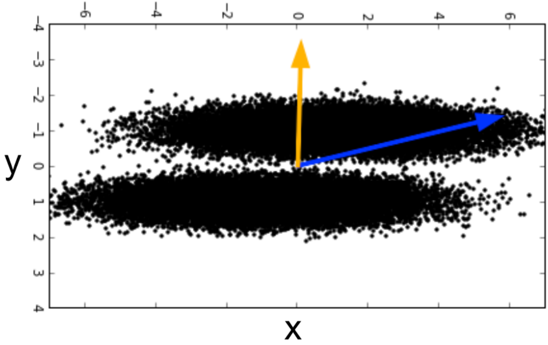
$$\tilde{\mathbf{E}}^\top = \tilde{\mathbf{U}}_d \quad \tilde{\mathbf{D}}^\top = \mathbf{\Sigma}_d \tilde{\mathbf{U}}_d^\top$$

and defining $\mathbf{U}_d = \mathbf{C}_{xx}^{-\frac{1}{2}} \tilde{\mathbf{U}}_d$ the non-whitened versions:

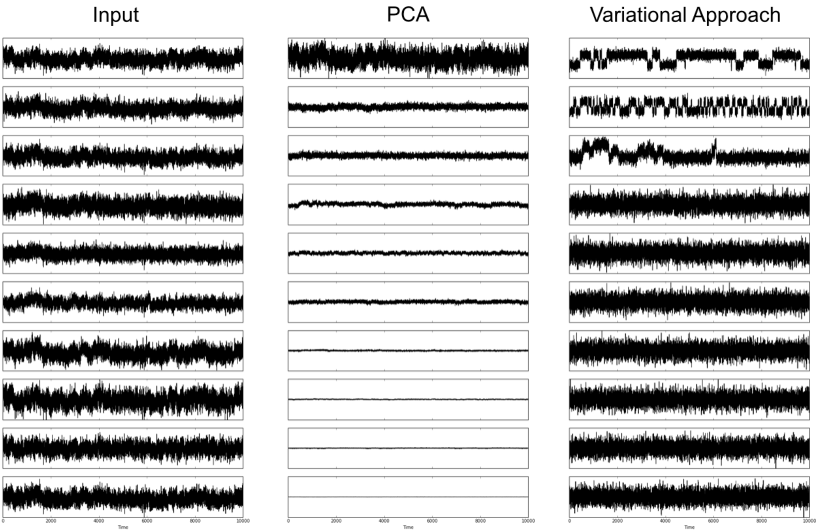
$$\mathbf{E}^\top = \mathbf{U}_d \quad \mathbf{D}^\top = \mathbf{\Sigma}_d \mathbf{U}_d^\top$$

- **Blind source separation**: Molgedey and Schuster. **Phys. Rev. Lett.** 72, 3634 (1994).
- **Time-lagged independent component analysis (TICA)**: G. Pérez-Hernández, F. Paul, T. Giorgino, G. De Fabritiis, F. Noé: Identification of slow molecular order parameters for Markov model construction. J. Chem. Phys. 139, 015102 (2013).

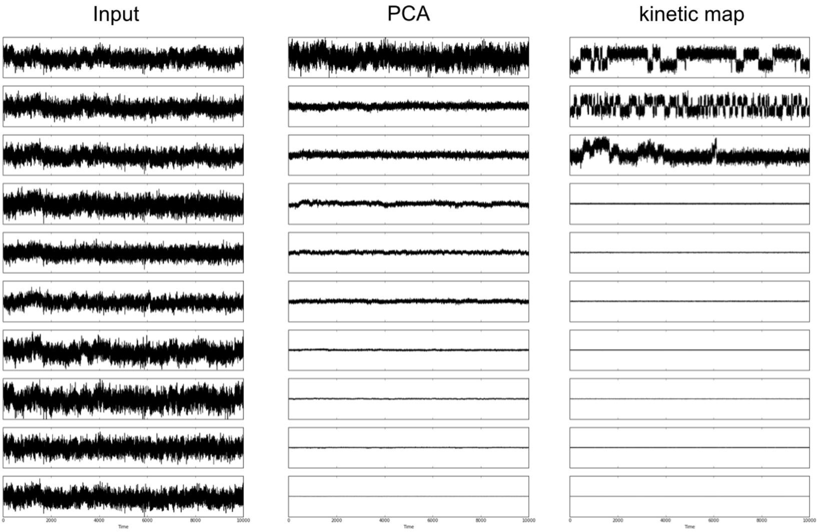
TICA/TCCA vs PCA



TICA/TCCA vs PCA

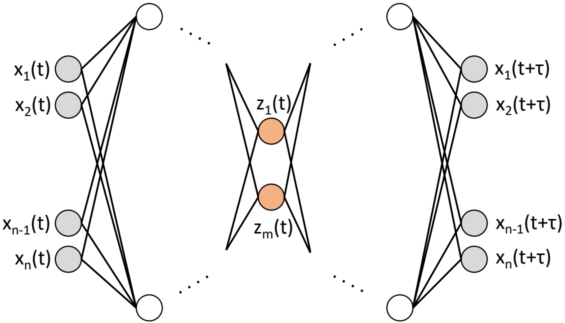


TICA/TCCA vs PCA



F. Noé, C. Clementi: Kinetic distance and kinetic maps from molecular dynamics simulation. J. Chem. Theory Comput. 11, 5002-5011 (2015)

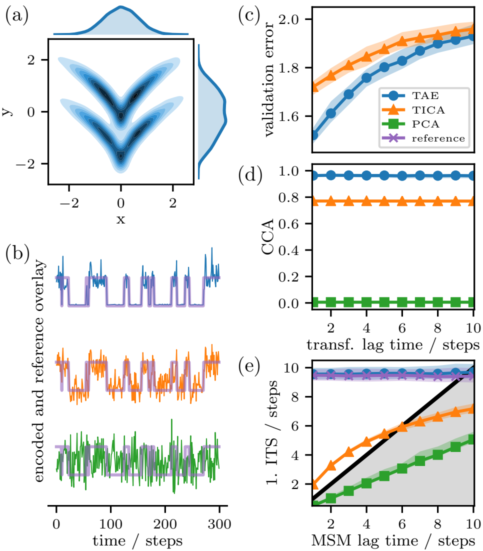
Time-Autoencoder Network



$$\min_{\theta_E,\theta_D} \|\mathbf{Y} - D(E(\mathbf{X};\theta_E);\theta_D)\|_F^2 = \min \sum_{t=1}^{T-\tau} \|\mathbf{x}_{t+\tau} - D(E(\mathbf{x}_t;\theta_E);\theta_D)\|^2$$

C. Wehmeyer and F. Noé: **Time-lagged autoencoders**: Deep learning of slow collective variables for molecular kinetics. J. Chem. Phys. 148, 241703 (2018)

Time-Autoencoder Network



TCCA maximizes correlation between vectors

- **Single collective coordinate**: TCCA maximizes

$$\tilde{\mathbf{u}}, \tilde{\mathbf{v}} = \arg\max_{\mathbf{u},\mathbf{v}} \frac{\mathbf{u}^\top \mathbf{C}_{xy} \mathbf{v}}{\sqrt{\mathbf{u}^\top \mathbf{C}_{xx} \mathbf{u}} \sqrt{\mathbf{v}^\top \mathbf{C}_{yy} \mathbf{v}}} = \arg\max_{\mathbf{u},\mathbf{v}} \frac{c_{\mathbf{uv}}(\tau)}{\sqrt{c_{\mathbf{uu}} c_{\mathbf{vv}}}}$$

- **Multiple collective coordinate** $\mathbf{u}_1, \ldots, \mathbf{u}_n$, $\mathbf{v}_1, \ldots, \mathbf{v}_n$:

$$\tilde{\mathbf{u}}_1, \ldots, \tilde{\mathbf{u}}_n, \tilde{\mathbf{v}}_1, \ldots, \tilde{\mathbf{v}}_n = \arg\max \sum_{i=1}^{n} \frac{c_{\mathbf{u}_1\mathbf{v}_1}(\tau)}{\sqrt{c_{\mathbf{u}_1\mathbf{u}_1} c_{\mathbf{v}_1\mathbf{v}_1}}}$$

$$s.t. \sum_t \mathbf{u}_i^\top \mathbf{x}_t \mathbf{v}_j^\top \mathbf{x}_{t+\tau} = \delta_{ij}$$

$$= \arg\max \left\| \mathbf{C}_{uu}^{-\frac{1}{2}} \mathbf{C}_{uv} \mathbf{C}_{vv}^{-\frac{1}{2}} \right\|_{tr}$$

$$= \arg\max \left\| \mathbf{C}_{uu}^{-\frac{1}{2}} \mathbf{C}_{uv} \mathbf{C}_{vv}^{-\frac{1}{2}} \right\|_{F}^{2}$$

- Trace norm and Frobenius norm:

$$\left\| \mathbf{C}_{uu}^{-\frac{1}{2}} \mathbf{C}_{uv} \mathbf{C}_{vv}^{-\frac{1}{2}} \right\|_{tr} = \sum_{i=1}^{n} \sigma_i \qquad \left\| \mathbf{C}_{uu}^{-\frac{1}{2}} \mathbf{C}_{uv} \mathbf{C}_{vv}^{-\frac{1}{2}} \right\|_{F}^{2} = \sum_{i=1}^{n} \sigma_i^2$$

with σ_i singular values of $\mathbf{C}_{uu}^{-\frac{1}{2}} \mathbf{C}_{uv} \mathbf{C}_{vv}^{-\frac{1}{2}}$.

TCCA maximizes correlation between vectors

- **Single collective coordinate**: TCCA maximizes

$$\tilde{\mathbf{u}}, \tilde{\mathbf{v}} = \arg\max_{\mathbf{u},\mathbf{v}} \frac{\mathbf{u}^\top \mathbf{C}_{xy} \mathbf{v}}{\sqrt{\mathbf{u}^\top \mathbf{C}_{xx} \mathbf{u}} \sqrt{\mathbf{v}^\top \mathbf{C}_{yy} \mathbf{v}}} = \arg\max_{\mathbf{u},\mathbf{v}} \frac{c_{\mathbf{uv}}(\tau)}{\sqrt{c_{\mathbf{uu}} c_{\mathbf{vv}}}}$$

- **Multiple collective coordinate $\mathbf{u}_1, \ldots, \mathbf{u}_n$, $\mathbf{v}_1, \ldots, \mathbf{v}_n$**:

$$\tilde{\mathbf{u}}_1, \ldots, \tilde{\mathbf{u}}_n, \tilde{\mathbf{v}}_1, \ldots, \tilde{\mathbf{v}}_n = \arg\max \sum_{i=1}^{n} \frac{c_{\mathbf{u}_1\mathbf{v}_1}(\tau)}{\sqrt{c_{\mathbf{u}_1\mathbf{u}_1} c_{\mathbf{v}_1\mathbf{v}_1}}}$$

$$s.t. \sum_t \mathbf{u}_i^\top \mathbf{x}_t \mathbf{v}_j^\top \mathbf{x}_{t+\tau} = \delta_{ij}$$

$$= \arg\max \left\| \mathbf{C}_{uu}^{-\frac{1}{2}} \mathbf{C}_{uv} \mathbf{C}_{vv}^{-\frac{1}{2}} \right\|_{tr}$$

$$= \arg\max \left\| \mathbf{C}_{uu}^{-\frac{1}{2}} \mathbf{C}_{uv} \mathbf{C}_{vv}^{-\frac{1}{2}} \right\|_{F}^{2}$$

- Trace norm and Frobenius norm:

$$\left\| \mathbf{C}_{uu}^{-\frac{1}{2}} \mathbf{C}_{uv} \mathbf{C}_{vv}^{-\frac{1}{2}} \right\|_{tr} = \sum_{i=1}^{n} \sigma_i \qquad \left\| \mathbf{C}_{uu}^{-\frac{1}{2}} \mathbf{C}_{uv} \mathbf{C}_{vv}^{-\frac{1}{2}} \right\|_{F}^{2} = \sum_{i=1}^{n} \sigma_i^2$$

with σ_i singular values of $\mathbf{C}_{uu}^{-\frac{1}{2}} \mathbf{C}_{uv} \mathbf{C}_{vv}^{-\frac{1}{2}}$.

TCCA maximizes correlation between vectors

- **Single collective coordinate**: TCCA maximizes

$$\tilde{\mathbf{u}}, \tilde{\mathbf{v}} = \arg\max_{\mathbf{u},\mathbf{v}} \frac{\mathbf{u}^\top \mathbf{C}_{xy} \mathbf{v}}{\sqrt{\mathbf{u}^\top \mathbf{C}_{xx} \mathbf{u}} \sqrt{\mathbf{v}^\top \mathbf{C}_{yy} \mathbf{v}}} = \arg\max_{\mathbf{u},\mathbf{v}} \frac{c_{\mathbf{uv}}(\tau)}{\sqrt{c_{\mathbf{uu}} c_{\mathbf{vv}}}}$$

- **Multiple collective coordinate $\mathbf{u}_1, ..., \mathbf{u}_n$, $\mathbf{v}_1, ..., \mathbf{v}_n$**:

$$\tilde{\mathbf{u}}_1, ..., \tilde{\mathbf{u}}_n, \tilde{\mathbf{v}}_1, ..., \tilde{\mathbf{v}}_n = \arg\max \sum_{i=1}^{n} \frac{c_{\mathbf{u}_1 \mathbf{v}_1}(\tau)}{\sqrt{c_{\mathbf{u}_1 \mathbf{u}_1} c_{\mathbf{v}_1 \mathbf{v}_1}}}$$

$$s.t. \sum_t \mathbf{u}_i^\top \mathbf{x}_t \mathbf{v}_j^\top \mathbf{x}_{t+\tau} = \delta_{ij}$$

$$= \arg\max \left\| \mathbf{C}_{uu}^{-\frac{1}{2}} \mathbf{C}_{uv} \mathbf{C}_{vv}^{-\frac{1}{2}} \right\|_{tr}$$

$$= \arg\max \left\| \mathbf{C}_{uu}^{-\frac{1}{2}} \mathbf{C}_{uv} \mathbf{C}_{vv}^{-\frac{1}{2}} \right\|_{F}^{2}$$

- Trace norm and Frobenius norm:

$$\left\| \mathbf{C}_{uu}^{-\frac{1}{2}} \mathbf{C}_{uv} \mathbf{C}_{vv}^{-\frac{1}{2}} \right\|_{tr} = \sum_{i=1}^{n} \sigma_i \qquad \left\| \mathbf{C}_{uu}^{-\frac{1}{2}} \mathbf{C}_{uv} \mathbf{C}_{vv}^{-\frac{1}{2}} \right\|_{F}^{2} = \sum_{i=1}^{n} \sigma_i^2$$

with σ_i singular values of $\mathbf{C}_{uu}^{-\frac{1}{2}} \mathbf{C}_{uv} \mathbf{C}_{vv}^{-\frac{1}{2}}$.

TICA maximizes autocovariances

- **Single collective coordinate**: TICA maximizes

$$\tilde{\mathbf{u}} = \arg\max_{\mathbf{u}} \frac{\mathbf{u}^\top \mathbf{C}_{xy}\mathbf{u}}{\sqrt{\mathbf{u}^\top \mathbf{C}_{xx}\mathbf{u}}} = \frac{c_{uu}(\tau)}{c_{uu}}$$

- Multiple directions $\tilde{\mathbf{u}}_1, \ldots, \tilde{\mathbf{u}}_n$

$$\tilde{\mathbf{u}}_1, \ldots, \tilde{\mathbf{u}}_n = \arg\max \sum_{i=1}^{n} \frac{c_{\mathbf{u}_1\mathbf{u}_1}(\tau)}{\sqrt{c_{\mathbf{u}_1\mathbf{u}_1} c_{\mathbf{v}_1\mathbf{v}_1}}}$$

$$s.t. \sum_t \mathbf{u}_i^\top \mathbf{x}_t \mathbf{u}_j^\top \mathbf{x}_{t+\tau} = \delta_{ij}$$

$$= \arg\max \left\| \mathbf{C}_{uu}^{-\frac{1}{2}} \mathbf{C}_{uv} \mathbf{C}_{vv}^{-\frac{1}{2}} \right\|_{tr} = \arg\max \left\| \mathbf{C}_{uu}^{-\frac{1}{2}} \mathbf{C}_{uv} \mathbf{C}_{vv}^{-\frac{1}{2}} \right\|_{F}^{2}$$

- TICA components are *white*:

$$\mathbf{Z}^\top \mathbf{Z} = \left(\mathbf{X}\mathbf{C}_{xx}^{-\frac{1}{2}} \tilde{\mathbf{U}}_d \right)^\top \mathbf{X}\mathbf{C}_{xx}^{-\frac{1}{2}} \tilde{\mathbf{U}}_d = \tilde{\mathbf{U}}_d^\top \tilde{\mathbf{U}}_d = \mathbf{I}$$

- The autocorrelations of IC's are identical to the Eigenvalues:

$$\left(\mathbf{X}\mathbf{C}_{xx}^{-\frac{1}{2}} \tilde{\mathbf{U}}_d \right)^\top \mathbf{Y}\mathbf{C}_{xx}^{-\frac{1}{2}} \tilde{\mathbf{U}}_d = \tilde{\mathbf{U}}_d^\top \mathbf{C}_{xx}^{-\frac{1}{2}} \mathbf{C}_{xy} \mathbf{C}_{xx}^{-\frac{1}{2}} \tilde{\mathbf{U}}_d = \mathbf{\Sigma}$$

TICA maximizes autocovariances

- **Single collective coordinate**: TICA maximizes

$$\tilde{\mathbf{u}} = \arg\max_{\mathbf{u}} \frac{\mathbf{u}^\top \mathbf{C}_{xy}\mathbf{u}}{\sqrt{\mathbf{u}^\top \mathbf{C}_{xx}\mathbf{u}}} = \frac{c_{uu}(\tau)}{c_{uu}}$$

- Multiple directions $\tilde{\mathbf{u}}_1, \dots, \tilde{\mathbf{u}}_n$

$$\tilde{\mathbf{u}}_1, \dots, \tilde{\mathbf{u}}_n = \arg\max \sum_{i=1}^{n} \frac{c_{\mathbf{u}_1\mathbf{u}_1}(\tau)}{\sqrt{c_{\mathbf{u}_1\mathbf{u}_1} c_{\mathbf{v}_1\mathbf{v}_1}}}$$

$$s.t. \sum_t \mathbf{u}_i^\top \mathbf{x}_t \mathbf{u}_j^\top \mathbf{x}_{t+\tau} = \delta_{ij}$$

$$= \arg\max \left\| \mathbf{C}_{uu}^{-\frac{1}{2}} \mathbf{C}_{uv} \mathbf{C}_{vv}^{-\frac{1}{2}} \right\|_{tr} = \arg\max \left\| \mathbf{C}_{uu}^{-\frac{1}{2}} \mathbf{C}_{uv} \mathbf{C}_{vv}^{-\frac{1}{2}} \right\|_F^2$$

- TICA components are *white*:

$$\mathbf{Z}^\top \mathbf{Z} = \left(\mathbf{X}\mathbf{C}_{xx}^{-\frac{1}{2}}\tilde{\mathbf{U}}_d\right)^\top \mathbf{X}\mathbf{C}_{xx}^{-\frac{1}{2}}\tilde{\mathbf{U}}_d = \tilde{\mathbf{U}}_d^\top \tilde{\mathbf{U}}_d = \mathbf{I}$$

- The autocorrelations of IC's are identical to the Eigenvalues:

$$\left(\mathbf{X}\mathbf{C}_{xx}^{-\frac{1}{2}}\tilde{\mathbf{U}}_d\right)^\top \mathbf{Y}\mathbf{C}_{xx}^{-\frac{1}{2}}\tilde{\mathbf{U}}_d = \tilde{\mathbf{U}}_d^\top \mathbf{C}_{xx}^{-\frac{1}{2}} \mathbf{C}_{xy} \mathbf{C}_{xx}^{-\frac{1}{2}} \tilde{\mathbf{U}}_d = \mathbf{\Sigma}$$

TICA maximizes autocovariances

- **Single collective coordinate**: TICA maximizes

$$\tilde{\mathbf{u}} = \arg\max_{\mathbf{u}} \frac{\mathbf{u}^\top \mathbf{C}_{xy}\mathbf{u}}{\sqrt{\mathbf{u}^\top \mathbf{C}_{xx}\mathbf{u}}} = \frac{c_{uu}(\tau)}{c_{uu}}$$

- Multiple directions $\tilde{\mathbf{u}}_1, ..., \tilde{\mathbf{u}}_n$

$$\tilde{\mathbf{u}}_1, ..., \tilde{\mathbf{u}}_n = \arg\max \sum_{i=1}^{n} \frac{c_{\mathbf{u}_1\mathbf{u}_1}(\tau)}{\sqrt{c_{\mathbf{u}_1\mathbf{u}_1} c_{\mathbf{v}_1\mathbf{v}_1}}}$$

$$s.t. \sum_t \mathbf{u}_i^\top \mathbf{x}_t \mathbf{u}_j^\top \mathbf{x}_{t+\tau} = \delta_{ij}$$

$$= \arg\max \left\| \mathbf{C}_{uu}^{-\frac{1}{2}} \mathbf{C}_{uv} \mathbf{C}_{vv}^{-\frac{1}{2}} \right\|_{tr} = \arg\max \left\| \mathbf{C}_{uu}^{-\frac{1}{2}} \mathbf{C}_{uv} \mathbf{C}_{vv}^{-\frac{1}{2}} \right\|_F^2$$

- TICA components are *white*:

$$\mathbf{Z}^\top \mathbf{Z} = \left(\mathbf{X}\mathbf{C}_{xx}^{-\frac{1}{2}} \tilde{\mathbf{U}}_d\right)^\top \mathbf{X}\mathbf{C}_{xx}^{-\frac{1}{2}} \tilde{\mathbf{U}}_d = \tilde{\mathbf{U}}_d^\top \tilde{\mathbf{U}}_d = \mathbf{I}$$

- The autocorrelations of IC's are identical to the Eigenvalues:

$$\left(\mathbf{X}\mathbf{C}_{xx}^{-\frac{1}{2}} \tilde{\mathbf{U}}_d\right)^\top \mathbf{Y}\mathbf{C}_{xx}^{-\frac{1}{2}} \tilde{\mathbf{U}}_d = \tilde{\mathbf{U}}_d^\top \mathbf{C}_{xx}^{-\frac{1}{2}} \mathbf{C}_{xy} \mathbf{C}_{xx}^{-\frac{1}{2}} \tilde{\mathbf{U}}_d = \mathbf{\Sigma}$$

TICA maximizes autocovariances

- **Single collective coordinate**: TICA maximizes

$$\tilde{\mathbf{u}} = \arg\max_{\mathbf{u}} \frac{\mathbf{u}^\top \mathbf{C}_{xy} \mathbf{u}}{\sqrt{\mathbf{u}^\top \mathbf{C}_{xx} \mathbf{u}}} = \frac{c_{uu}(\tau)}{c_{uu}}$$

- Multiple directions $\tilde{\mathbf{u}}_1, \ldots, \tilde{\mathbf{u}}_n$

$$\tilde{\mathbf{u}}_1, \ldots, \tilde{\mathbf{u}}_n = \arg\max \sum_{i=1}^{n} \frac{c_{\mathbf{u}_1\mathbf{u}_1}(\tau)}{\sqrt{c_{\mathbf{u}_1\mathbf{u}_1} c_{\mathbf{v}_1\mathbf{v}_1}}}$$

$$s.t. \sum_t \mathbf{u}_i^\top \mathbf{x}_t \mathbf{u}_j^\top \mathbf{x}_{t+\tau} = \delta_{ij}$$

$$= \arg\max \left\| \mathbf{C}_{uu}^{-\frac{1}{2}} \mathbf{C}_{uv} \mathbf{C}_{vv}^{-\frac{1}{2}} \right\|_{tr} = \arg\max \left\| \mathbf{C}_{uu}^{-\frac{1}{2}} \mathbf{C}_{uv} \mathbf{C}_{vv}^{-\frac{1}{2}} \right\|_F^2$$

- TICA components are *white*:

$$\mathbf{Z}^\top \mathbf{Z} = \left(\mathbf{X} \mathbf{C}_{xx}^{-\frac{1}{2}} \tilde{\mathbf{U}}_d \right)^\top \mathbf{X} \mathbf{C}_{xx}^{-\frac{1}{2}} \tilde{\mathbf{U}}_d = \tilde{\mathbf{U}}_d^\top \tilde{\mathbf{U}}_d = \mathbf{I}$$

- The autocorrelations of IC's are identical to the Eigenvalues:

$$\left(\mathbf{X} \mathbf{C}_{xx}^{-\frac{1}{2}} \tilde{\mathbf{U}}_d \right)^\top \mathbf{Y} \mathbf{C}_{xx}^{-\frac{1}{2}} \tilde{\mathbf{U}}_d = \tilde{\mathbf{U}}_d^\top \mathbf{C}_{xx}^{-\frac{1}{2}} \mathbf{C}_{xy} \mathbf{C}_{xx}^{-\frac{1}{2}} \tilde{\mathbf{U}}_d = \mathbf{\Sigma}$$

Special case: Markov State Model

- Choose characteristic functions

$$1_i(\mathbf{x}) = \begin{cases} 1 & \mathbf{x} \in S_i \\ 0 & \mathbf{x} \neq S_i \end{cases}$$

with sets $S_1, ..., S_n$ that partition the state space $\Omega = \cup_i S_i$

- Covariance matrices using $\pi_i = \mathbb{P}[\mathbf{x} \in S_i]$:

$$(\mathbf{C}_{xx})_{ij} = \mathbb{E}_t \left[1_i(\mathbf{x}_t) \cdot 1_j(\mathbf{x}_t)\right] = \delta_{ij}\pi_i$$
$$\mathbf{C}_{xx} = \mathbf{\Pi}$$

- And

$$(\mathbf{C}_{xy})_{ij} = \mathbb{E}\left[1_i(\mathbf{x}_t) \cdot 1_j(\mathbf{x}_{t+\tau})\right] = \pi_i p_{ij}$$

with $p_{ij} = \mathbb{P}\left[\mathbf{x}_{t+\tau} \in S_j \mid \mathbf{x}_t \in S_i\right]$.

- Thus we get an ordinary eigenvalue problem with the transition matrix:

$$\mathbf{C}_{xx}^{-\frac{1}{2}} \mathbf{C}_{xy} \mathbf{C}_{xx}^{-\frac{1}{2}} \mathbf{U} = \mathbf{U}\mathbf{\Lambda}$$
$$\mathbf{P}\mathbf{\Pi}^{-\frac{1}{2}} \tilde{\mathbf{U}} = \mathbf{\Pi}^{-\frac{1}{2}} \tilde{\mathbf{U}}\mathbf{\Lambda}$$
$$\mathbf{P}\mathbf{U} = \mathbf{U}\mathbf{\Lambda}$$

Special case: Markov State Model

- Choose characteristic functions

$$1_i(\mathbf{x}) = \begin{cases} 1 & \mathbf{x} \in S_i \\ 0 & \mathbf{x} \neq S_i \end{cases}$$

  with sets $S_1, ..., S_n$ that partition the state space $\Omega = \cup_i S_i$
- Covariance matrices using $\pi_i = \mathbb{P}[\mathbf{x} \in S_i]$:

$$(\mathbf{C}_{xx})_{ij} = \mathbb{E}_t\left[1_i(\mathbf{x}_t) \cdot 1_j(\mathbf{x}_t)\right] = \delta_{ij}\pi_i$$
$$\mathbf{C}_{xx} = \mathbf{\Pi}$$

- And

$$(\mathbf{C}_{xy})_{ij} = \mathbb{E}\left[1_i(\mathbf{x}_t) \cdot 1_j(\mathbf{x}_{t+\tau})\right] = \pi_i p_{ij}$$

  with $p_{ij} = \mathbb{P}\left[\mathbf{x}_{t+\tau} \in S_j \mid \mathbf{x}_t \in S_i\right]$.
- Thus we get an ordinary eigenvalue problem with the transition matrix:

$$\mathbf{C}_{xx}^{-\frac{1}{2}}\mathbf{C}_{xy}\mathbf{C}_{xx}^{-\frac{1}{2}}\mathbf{U} = \mathbf{U}\mathbf{\Lambda}$$
$$\mathbf{P}\mathbf{\Pi}^{-\frac{1}{2}}\tilde{\mathbf{U}} = \mathbf{\Pi}^{-\frac{1}{2}}\tilde{\mathbf{U}}\mathbf{\Lambda}$$
$$\mathbf{P}\mathbf{U} = \mathbf{U}\mathbf{\Lambda}$$

Special case: Markov State Model

- Choose characteristic functions

$$1_i(\mathbf{x}) = \begin{cases} 1 & \mathbf{x} \in S_i \\ 0 & \mathbf{x} \neq S_i \end{cases}$$

  with sets $S_1, ..., S_n$ that partition the state space $\Omega = \cup_i S_i$
- Covariance matrices using $\pi_i = \mathbb{P}[\mathbf{x} \in S_i]$:

$$(\mathbf{C}_{xx})_{ij} = \mathbb{E}_t\left[1_i(\mathbf{x}_t) \cdot 1_j(\mathbf{x}_t)\right] = \delta_{ij}\pi_i$$
$$\mathbf{C}_{xx} = \mathbf{\Pi}$$

- And

$$(\mathbf{C}_{xy})_{ij} = \mathbb{E}\left[1_i(\mathbf{x}_t) \cdot 1_j(\mathbf{x}_{t+\tau})\right] = \pi_i p_{ij}$$

  with $p_{ij} = \mathbb{P}\left[\mathbf{x}_{t+\tau} \in S_j \mid \mathbf{x}_t \in S_i\right]$.
- Thus we get an ordinary eigenvalue problem with the transition matrix:

$$\mathbf{C}_{xx}^{-\frac{1}{2}} \mathbf{C}_{xy} \mathbf{C}_{xx}^{-\frac{1}{2}} \mathbf{U} = \mathbf{U}\mathbf{\Lambda}$$
$$\mathbf{P}\mathbf{\Pi}^{-\frac{1}{2}} \tilde{\mathbf{U}} = \mathbf{\Pi}^{-\frac{1}{2}} \tilde{\mathbf{U}}\mathbf{\Lambda}$$
$$\mathbf{P}\mathbf{U} = \mathbf{U}\mathbf{\Lambda}$$

Special case: Markov State Model

- Choose characteristic functions

$$1_i(\mathbf{x}) = \begin{cases} 1 & \mathbf{x} \in S_i \\ 0 & \mathbf{x} \neq S_i \end{cases}$$

  with sets $S_1, ..., S_n$ that partition the state space $\Omega = \cup_i S_i$
- Covariance matrices using $\pi_i = \mathbb{P}[\mathbf{x} \in S_i]$:

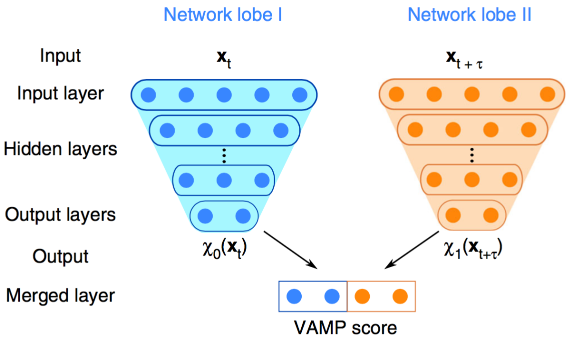
$$(\mathbf{C}_{xx})_{ij} = \mathbb{E}_t \left[1_i(\mathbf{x}_t) \cdot 1_j(\mathbf{x}_t)\right] = \delta_{ij}\pi_i$$
$$\mathbf{C}_{xx} = \mathbf{\Pi}$$

- And

$$(\mathbf{C}_{xy})_{ij} = \mathbb{E}\left[1_i(\mathbf{x}_t) \cdot 1_j(\mathbf{x}_{t+\tau})\right] = \pi_i p_{ij}$$

  with $p_{ij} = \mathbb{P}\left[\mathbf{x}_{t+\tau} \in S_j \mid \mathbf{x}_t \in S_i\right].$
- Thus we get an ordinary eigenvalue problem with the transition matrix:

$$\mathbf{C}_{xx}^{-\frac{1}{2}} \mathbf{C}_{xy} \mathbf{C}_{xx}^{-\frac{1}{2}} \mathbf{U} = \mathbf{U}\mathbf{\Lambda}$$
$$\mathbf{P}\mathbf{\Pi}^{-\frac{1}{2}} \tilde{\mathbf{U}} = \mathbf{\Pi}^{-\frac{1}{2}} \tilde{\mathbf{U}}\mathbf{\Lambda}$$
$$\mathbf{P}\mathbf{U} = \mathbf{U}\mathbf{\Lambda}$$



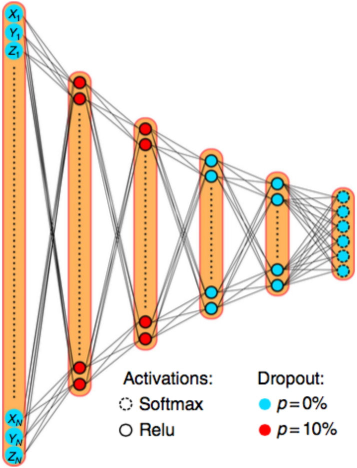
Objective: maximize VAMP-2 score.

$$\arg\max_{\chi_0,\chi_1} \left\| \mathbf{C}_{00}^{-\frac{1}{2}} \mathbf{C}_{01} \mathbf{C}_{11}^{-\frac{1}{2}} \right\|_F^2$$

with covariance matrices (can be computed in minibatches):

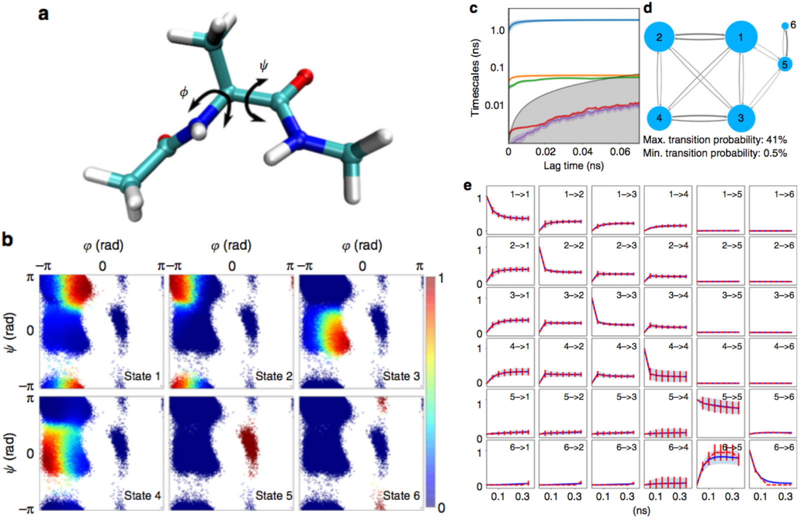
$$\mathbf{C}_{00} = \sum_t \chi_0(\mathbf{x}_t)\chi_0^\top(\mathbf{x}_t) \quad \mathbf{C}_{01} = \sum_t \chi_0(\mathbf{x}_t)\chi_1^\top(\mathbf{x}_{t+\tau}) \quad \mathbf{C}_{11} = \sum_t \chi_1(\mathbf{x}_{t+\tau})\chi_1^\top(\mathbf{x}_{t+\tau}).$$

VAMPnet



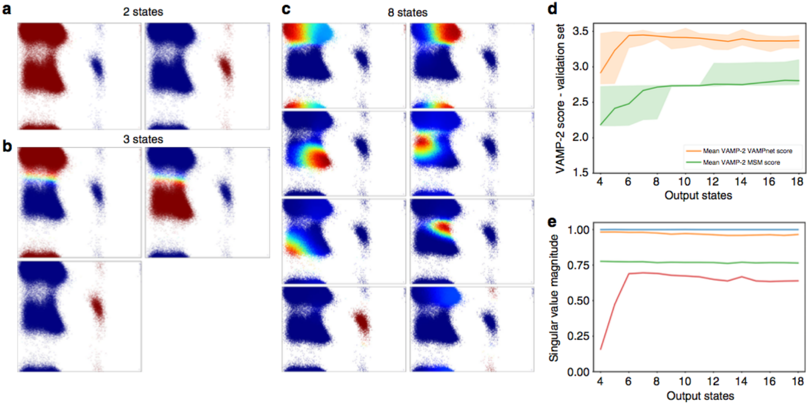
A. Mardt, L. Pasquali, H. Wu, F. Noé: **VAMPnets** for deep learning of molecular kinetics Nat. Commun. 9, 5 (2018)

VAMPnet



A. Mardt, L. Pasquali, H. Wu, F. Noé: **VAMPnets** for deep learning of molecular kinetics Nat. Commun. 9, 5 (2018)

VAMPnet



A. Mardt, L. Pasquali, H. Wu, F. Noé: **VAMPnets** for deep learning of molecular kinetics Nat. Commun. 9, 5 (2018)