

Recall, for some trajectory  $|X_t\rangle$ , the time lag correlation matrix ( $C^{(\Delta t)}$ ) and covariance matrix are defined as:

$$C^{(\Delta t)} = \mathbb{E}[|X_t\rangle\langle X_{t+\Delta t}|]$$

$$\Sigma = \mathbb{E}[|X_t\rangle\langle X_t|]$$

We expect that  $C^{(\Delta t)}$  is symmetric, and so one trivial way to ensure that it is, is to calculate the sample mean as stated above, and symmetrize this estimate:

$$C^{(\Delta t)} = \frac{1}{2(N - \Delta t)} \sum_{t=1}^{N-\Delta t} |X_t\rangle\langle X_{t+\Delta t}| + |X_{t+\Delta t}\rangle\langle X_t|$$

while the covariance matrix is estimated as the sample covariance matrix. If you instead try to be a bit more rigorous you can prove that the above estimate is an MLE, but this changes the estimate of  $\Sigma$  as well! So, if you assume that the  $|X_t\rangle$  and  $|X_{t+\Delta t}\rangle$  are separate random variables and are distributed according to a multivariate normal. The MLE estimators of the two matrices are:

$$C_{mle}^{(\Delta t)} = \frac{1}{2(N - \Delta t)} \sum_{t=1}^{N-\Delta t} |X_t\rangle\langle X_{t+\Delta t}| + |X_{t+\Delta t}\rangle\langle X_t|$$

$$\Sigma_{mle} = \frac{1}{2(N - \Delta t)} \sum_{t=1}^{N-\Delta t} |X_t\rangle\langle X_t| + |X_{t+\Delta t}\rangle\langle X_{t+\Delta t}|$$

These two solutions correspond to the two branches: tICA and tica\_mle in my msmbuilder fork.