Supplementary Information for “Modeling Molecular Kinetics with tICA and the Kernel Trick”

Christian R. Schwantes† and Vijay S. Pande*,†,§

Department of Chemistry, Stanford University, Stanford, CA 94305

E-mail: pande@stanford.edu

*To whom correspondence should be addressed
†Department of Chemistry, Stanford University, Stanford, CA 94305
‡Department of Computer Science, Stanford University, Stanford, CA 94305
§Department of Structural Biology, Stanford University, Stanford, CA 94305
Braket Notation

In the main text, we used
\[ x \cdot y = x^T y \]
to denote the inner product between two column vectors, and
\[ x \otimes y = xy^T \]
to denote the outer product. However, in what follows, it is very useful to use bra-ket notation to follow the algebraic steps that arrive at the ktICA solution. Therefore, below:
\[ \langle x | y \rangle = x \cdot y = x^T y \]
denotes the inner product, while the outer product is written as:
\[ |x \rangle \langle y | = x \otimes y = xy^T \]

Maximum Likelihood Estimator for tICA Matrices

If \( |x_t \rangle \) is a Markov chain in phase space, then the time-lag correlation, \( C^{(\tau)} \), and covariance, \( \Sigma \), matrices are defined as:
\[ C^{(\tau)} = \mathbb{E} \left[ |\delta x_t \rangle \langle \delta x_{t+\tau} | \right] \]  \hspace{1cm} (1)
\[ \Sigma = \mathbb{E} \left[ |\delta x_t \rangle \langle \delta x_t | \right] \]  \hspace{1cm} (2)
where \( |\delta x_t \rangle = |x_t \rangle - |\mu \rangle \) and \( \mu = \mathbb{E} \left[ |x_t \rangle \right] \).

To use the tICA method, we must construct estimators for \( \mu \), \( \Sigma \), and \( C^{(\tau)} \) given finite samples of the Markov chain. Importantly, the time-lag correlation matrix should be symmetric since the dynamics are reversible, but this may not be the case if only a sample mean
is used. The simplest approach we can take is to use a maximum likelihood estimator, where we assume the data is distributed according to a multivariate normal distribution.

We assume that we are given \( M \) pairs of transitions separated in time by \( \tau \), \( \{(|X_t\rangle, |Y_t\rangle)\}_{t=1}^{M} \). Define a new variable, \(|Z_t\rangle\), which is the concatenation of \(|X_t\rangle\) and \(|Y_t\rangle\):

\[
|Z_t\rangle = \begin{bmatrix} |X_t\rangle \\ |Y_t\rangle \end{bmatrix}
\]

Then we will assume that these variables are distributed according to a multivariate normal with covariance matrix, \( S \) equal to:

\[
S = \begin{bmatrix} \Sigma & C(\tau) \\ C(\tau) & \Sigma \end{bmatrix}
\]

and mean given by:

\[
|m\rangle = \begin{bmatrix} |\mu\rangle \\ |\mu\rangle \end{bmatrix}
\]

Then the probability density at \(|z\rangle\) can be written as:

\[
p(|z\rangle| S, \mu) = (2\pi)^{-\frac{d}{2}} |S|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \left( |z\rangle - |\langle m\rangle| \right) S^{-1} \left( |z\rangle - |\langle m\rangle| \right) \right]
\]

where \( d \) is twice the dimension of phase space.

Using this distribution, we can write the log-likelihood of the observed transitions, given the model:

\[
\log L = \sum_{t=1}^{M} \left[ -\frac{d}{2} \log 2\pi - \frac{1}{2} \log |S| - \frac{1}{2} \left( \langle Z_t\rangle - \langle m\rangle \right) S^{-1} \left( |Z_t\rangle - |\langle m\rangle| \right) \right]
\]

(3)

Using the properties of the matrix differential, it can be shown that the total differential of the log-likelihood is:
\[ d \log L = - \frac{M}{2} \text{tr}(S^{-1}dS) - \frac{1}{2} \sum_{t=1}^{M} \text{tr} \left[ 2S^{-1}\left( |Z_t\rangle - |m\rangle \right) d\langle m| \right] - \frac{1}{2} \sum_{t=1}^{M} \text{tr} \left[ S^{-1}\left( |Z_t\rangle - |m\rangle \right) \left( \langle Z_t| - \langle m| \right) S^{-1}dS \right] \] 

(4)

We could rewrite this total differential in terms of \( |\mu\rangle \), \( C^{(r)} \), and \( \Sigma \), but it’s more convenient to use the method of Lagrange multipliers to constrain the solutions, \( S \) and \( |m\rangle \). Let \( R \) be a block rotation matrix:

\[
R = \begin{bmatrix}
0 & I \\
I & 0
\end{bmatrix}
\]

It’s easy to show that:

\[
R \begin{bmatrix} A & B \\ C & D \end{bmatrix} R = \begin{bmatrix} D & C \\ B & A \end{bmatrix}
\]

and

\[
R \begin{bmatrix} A \\ B \end{bmatrix} = \begin{bmatrix} B \\ A \end{bmatrix},
\]

therefore the constraints we need to impose are:

\[
RSR = S \quad \text{and} \quad R|m\rangle = |m\rangle.
\]

We can then construct the Lagrange function:

\[
\Lambda = \log L + \langle \lambda | \left( R|m\rangle - |m\rangle \right) + \sum_i \sum_j \phi_{ij} [RSR - S]_{ij}
\]

The total differential of \( \Lambda \) can then be written:
\[ d\Lambda = \text{tr} \left[ \left( - \sum_{t=1}^{M} S^{-1}(|Z_t| - |m\rangle) + (R - I) |\lambda\rangle \right) d|m\rangle \right] \\
+ \text{tr} \left[ \left( -\frac{M}{2} S^{-1} - \frac{1}{2} \sum_{t=1}^{M} S^{-1}(|Z_t| - |m\rangle) \left( \langle Z_t| - \langle m| \right) S^{-1} + (R \Phi R - \Phi) \right) dS \right] \\
+ \text{tr} \left[ \left( R |m\rangle - |m\rangle \right) d\langle \lambda| \right] + \text{tr} \left[ \left( RSR - S \right) d\Phi \right] \\
\]

(5)

The total differential is zero exactly when the terms multiplying \(dS\) and \(d|m\rangle\) are zero. First, we solve for \( |m\rangle\):

\[ - \sum_{t=1}^{M} S^{-1}(|Z_t| - |m\rangle) + (R - I) |\lambda\rangle = 0 \]
\[ \left( \sum_{t=1}^{M} |Z_t\rangle \right) - M |m\rangle = S(R - I) |\lambda\rangle \]

Now, we add this equation to itself, but multiplied (from the left) by \(R\):

\[ \left( \sum_{t=1}^{M} |Z_t\rangle \right) - M |m\rangle + R \left( \sum_{t=1}^{M} |Z_t\rangle \right) - M |m\rangle = \left[ S(R - I) |\lambda\rangle \right] + R \left[ S(R - I) |\lambda\rangle \right] \\
\left( \sum_{t=1}^{M} |Z_t\rangle + R |Z_t\rangle \right) - 2M |m\rangle = \left( SR - S + RSR - RS \right) |\lambda\rangle \\
\left( \sum_{t=1}^{M} |Z_t\rangle + R |Z_t\rangle \right) - 2M |m\rangle = \left[ (SR - RS) + (RSR - S) \right] |\lambda\rangle \\
\left( \sum_{t=1}^{M} |Z_t\rangle + R |Z_t\rangle \right) - 2M |m\rangle = 0 \\
|m\rangle = \frac{1}{2M} \sum_{t=1}^{M} |Z_t\rangle + R |Z_t\rangle \]
We can then solve for $S$.

\[-\frac{M}{2}S^{-1} - \frac{1}{2} \sum_{t=1}^{M} S^{-1} |\delta Z_t\rangle \langle \delta Z_t| S^{-1} + (R\Phi R - \Phi) = 0\]

\[\frac{M}{2}S + \frac{1}{2} \sum_{t=1}^{M} |\delta Z_t\rangle \langle \delta Z_t| = S(R\Phi R - \Phi)S\]

Now, add this equation to itself, but multiplied by $R$ from the left and the right.

\[\frac{M}{2}S + \frac{1}{2} \sum_{t=1}^{M} |\delta Z_t\rangle \langle \delta Z_t| + R\left(\frac{M}{2}S + \sum_{t=1}^{M} |\delta Z_t\rangle \langle \delta Z_t|\right)R\]

\[= S(R\Phi R - \Phi)S + R\left(S(R\Phi R - \Phi)S\right)R\]

\[MS + \left(\frac{1}{2} \sum_{t=1}^{M} |\delta Z_t\rangle \langle \delta Z_t| + R|\delta Z_t\rangle \langle \delta Z_t| R\right)\]

\[= SR\Phi RS - S\Phi S + RSR\Phi RSR - RS\Phi SR\]

\[MS + \left(\frac{1}{2} \sum_{t=1}^{M} |\delta Z_t\rangle \langle \delta Z_t| + R|\delta Z_t\rangle \langle \delta Z_t| R\right)\]

\[= \left(SR\Phi RS - R\Phi S\right) + \left(RSR\Phi RSR - S\Phi S\right)\]

\[MS + \left(\frac{1}{2} \sum_{t=1}^{M} |\delta Z_t\rangle \langle \delta Z_t| + R|\delta Z_t\rangle \langle \delta Z_t| R\right)\]

\[= 0\]

\[S = \frac{1}{2M} \sum_{t=1}^{M} \left(|\delta Z_t\rangle \langle \delta Z_t| + R|\delta Z_t\rangle \langle \delta Z_t| R\right)\]

These solutions mean that the maximum likelihood estimators for $|\mu\rangle$, $\Sigma$, and $C^{(r)}$ are:

\[|\mu_{mle}\rangle = \frac{1}{2N} \sum_{t=1}^{N} \left(|\delta X_{t}\rangle + |\delta Y_{t}\rangle\right)\] (6)

\[\Sigma_{mle} = \frac{1}{2N} \sum_{t=1}^{N} \left(|\delta X_{t}\rangle \langle \delta X_{t}| + |\delta Y_{t}\rangle \langle \delta Y_{t}|\right)\] (7)
\[
C_{\text{mle}}^{(\tau)} = \frac{1}{2N} \sum_{t=1}^{N} \left( |\delta X_t\rangle \langle \delta Y_t| + |\delta Y_t\rangle \langle \delta X_t| \right)
\]  

(8)

Although, the MVN assumption is very crude, these estimators have two desirable properties:

1. \(C^{(\tau)}\) is always symmetric. Since the dynamics are reversible, the true time-lag correlation matrix is symmetric.

2. The Rayleigh quotient:

\[
\frac{\langle v | C_{\text{mle}}^{(\tau)} | v \rangle}{\langle v | \Sigma_{\text{mle}} | v \rangle}
\]

is always in \([-1, 1]\) (as long as \(\Sigma_{\text{mle}}\) is positive definite), which ensures that the eigenvalues from tICA are always real, and can be interpreted as timescales. This is because:

\[
\left| \frac{1}{2M} \sum_{t=1}^{M} \langle v | \delta X_t \rangle \langle \delta Y_t| v \rangle + \langle v | \delta Y_t \rangle \langle \delta X_t| v \rangle \right| \leq \left| \frac{1}{2M} \sum_{t=1}^{M} \langle v | \delta X_t \rangle \langle \delta X_t| v \rangle + \langle v | \delta Y_t \rangle \langle \delta Y_t| v \rangle \right|
\]

which follows from the Cauchy-Schwarz inequality.

**tICA Solutions are in the Span of the Input Data**

The solutions to the tICA problem satisfy:

\[
C^{(\tau)} |v\rangle = \lambda \Sigma |v\rangle
\]
Since $\Sigma$ is positive definite, it is also nonsingular and so we can write:

$$|v\rangle = \frac{1}{\lambda} \Sigma^{-1} C^{(r)} |v\rangle$$

$$= \frac{1}{2M\lambda} \Sigma^{-1} \sum_{t=1}^{M} \left( \langle \Phi(X_t) | v \rangle \right) |\Phi(Y_t)\rangle + \left( \langle \Phi(Y_t) | v \rangle \right) |\Phi(X_t)\rangle$$

$$:= \frac{1}{2M\lambda} \Sigma^{-1} |x\rangle$$

where we’ve defined $|x\rangle$ to be the sum from the equation above. We know the covariance matrix can also be diagonalized by a unitary matrix, $P$:

$$\Sigma = P \Lambda P^T = \begin{bmatrix} |p_1\rangle & |p_2\rangle & \cdots & |p_d\rangle \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_d \end{bmatrix} \begin{bmatrix} \langle p_1 | \\ \langle p_2 | \\ \vdots \\ \langle p_d | \end{bmatrix}$$

where $|p_i\rangle$ is an eigenvector of $\Sigma$ and $\lambda_i$ is its eigenvalue. It’s easy to show that the eigenvectors of $\Sigma$ are in the span of the $|X_t\rangle$’s and $|Y_t\rangle$’s. Using the decomposition above, we can rewrite the tICA solution in terms of a linear combination of the eigenvectors of $\Sigma$:

$$|v\rangle = \frac{1}{2M\lambda} P \Lambda^{-1} P^T |x\rangle$$

$$= \frac{1}{2M\lambda} \left[ |p_1\rangle |p_2\rangle \cdots |p_d\rangle \right] \Lambda^{-1} \begin{bmatrix} \langle p_1 | \\ \langle p_2 | \\ \vdots \\ \langle p_d | \end{bmatrix}$$

$$= \frac{1}{2M\lambda} \sum_{i=1}^{d} \left( \lambda_i^{-1} \langle p_i | x \rangle \right) |p_i\rangle$$
which means that $|v\rangle$ is in the span of the eigenvectors of $\Sigma$. Since the eigenvectors of $\Sigma$ are all in the span of the data, $|v\rangle$ is also in the span of the $|X_t\rangle$’s and $|Y_t\rangle$’s.

**Derivation of the ktICA Solution**

From the main text, recall that we are trying to rewrite the numerator and denominator of the tICA objective function in Eq. (9).

$$f(|v\rangle) = \frac{\langle v| C^{(\tau)} |v\rangle}{\langle v| \Sigma |v\rangle} \quad (9)$$

As shown above, the solution $|v\rangle$ is in the span of the input data, so let $\beta$ be the length $2M$ vector of coefficients such that:

$$|v\rangle = \sum_{t=1}^{M} \beta_t |\Phi(X_t)) + \beta_{i+M} |\Phi(Y_t)) \quad (10)$$
Now, we need to simply expand the numerator and denominator of Eq. (9) in terms of the elements of $\beta$.

\[
\langle v| C^{(\tau)} |v \rangle = \frac{1}{2M} \sum_{t=1}^{M} \langle v| \Phi(X_t) \rangle \langle \Phi(Y_t)| v \rangle \\
\quad + \langle v| \Phi(Y_t) \rangle \langle \Phi(X_t)| v \rangle
\]

\[
= \frac{1}{2M} \sum_{t=1}^{M} \left( \sum_{i=1}^{M} \left[ \beta_i K_{it}^{XX} + \beta_{i+M} K_{it}^{YX} \right] \sum_{j=1}^{M} \left[ \beta_j K_{tj}^{YX} + \beta_{j+M} K_{tj}^{YY} \right] + \sum_{i=1}^{M} \left[ \beta_i K_{it}^{XY} + \beta_{i+M} K_{it}^{XY} \right] \sum_{j=1}^{M} \left[ \beta_j K_{tj}^{XX} + \beta_{j+M} K_{tj}^{YY} \right] \right)
\]

\[
= \frac{1}{2M} \sum_{t=1}^{M} \left[ \beta^T \begin{pmatrix} K^{XX} \\ K^{XY} \end{pmatrix} \right] \left[ \begin{pmatrix} K^{YX} & K^{YY} \end{pmatrix} \beta \right]_t \\
\quad + \left[ \beta^T \begin{pmatrix} K^{XY} \\ K^{YY} \end{pmatrix} \right] \left[ \begin{pmatrix} K^{XX} & K^{XY} \end{pmatrix} \beta \right]_t
\]

\[
= \frac{1}{2M} \beta^T \begin{pmatrix} K^{XX} & K^{XY} \\ K^{YX} & K^{YY} \end{pmatrix} \begin{pmatrix} K^{YX} & K^{YY} \\ K^{XX} & K^{XY} \end{pmatrix} \beta
\]

\[
= \frac{1}{2M} \beta^T \begin{pmatrix} K^{XX}K^{YX} + K^{XY}K^{XX} & K^{XX}K^{YY} + K^{XY}K^{XY} \\ K^{YX}K^{YX} + K^{YY}K^{XX} & K^{YX}K^{YY} + K^{YY}K^{XY} \end{pmatrix} \beta
\]

\[
= \frac{1}{2M} \beta^T \begin{pmatrix} K^{XX} & K^{XY} \\ K^{YX} & K^{YY} \end{pmatrix} \begin{pmatrix} K^{YX} & K^{YY} \\ K^{XX} & K^{XY} \end{pmatrix} \beta
\]

\[
= \frac{1}{2M} \beta^T \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix} K \beta
\]
Let \( R = \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix} \), then the numerator becomes:

\[
\langle v \mid C^{(\tau)} \mid v \rangle = \frac{\beta^T K R K \beta}{2M} \tag{11}
\]

Through very analogous steps, it is easy to show that the denominator becomes:

\[
\langle v \mid \Sigma \mid v \rangle = \frac{\beta^T K K \beta}{2M} \tag{12}
\]

This means that the tICA method can be rewritten in terms of solely inner-products and we can use the kernel trick.

**Centering Data in the Feature Space**

In the proof of the ktICA solution, we assumed that the vectors, \(|\Phi(X_t)\rangle\), were centered (i.e. \( \mathbb{E}[|\Phi(X_t)\rangle] = 0 \)). In order to solve the tICA problem, we need to calculate the gram matrix, \( K \), between the centered points in \( V \). However, it is easy to show that the centered gram matrix can be calculated from the uncentered one:

\[
K = K_u - \frac{1}{2M} K_u 1 - \frac{1}{2M} 1 K_u + \frac{1}{4M^2} 1 K_u 1 \tag{13}
\]

where \( 1 \) is a \( 2M \times 2M \) matrix of all ones, and \( K_u \) is the gram matrix defined in the main text for the uncentered data.