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

$$\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^T \mathbf{y}$$

to denote the inner product between two column vectors, and

$$\mathbf{x} \otimes \mathbf{y} = \mathbf{x}\mathbf{y}^T$$

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:

$$\langle \mathbf{x} | \, \mathbf{y} \rangle = \mathbf{x} \cdot \mathbf{y} = \mathbf{x}^T \mathbf{y}$$

denotes the inner product, while the outer product is written as:

$$|\mathbf{x}\rangle \langle \mathbf{y}| = \mathbf{x} \otimes \mathbf{y} = \mathbf{x}\mathbf{y}^T$$

Maximum Likelihood Estimator for tICA Matrices

If $|\mathbf{x}_t\rangle$ is a Markov chain in phase space, then the time-lag correlation, $C^{(\tau)}$, and covariance, Σ, matrices are defined as:

$$C^{(\tau)} = \mathbb{E}\big[\, |\delta \mathbf{x}_t\rangle \langle \delta \mathbf{x}_{t+\tau}| \, \big] \tag{1}$$

$$\Sigma = \mathbb{E}\big[\, |\delta \mathbf{x}_t\rangle \langle \delta \mathbf{x}_t| \, \big] \tag{2}$$

where $|\delta \mathbf{x}_t\rangle = |\mathbf{x}_t\rangle - |\mu\rangle$ and $\mu = \mathbb{E}\big[\, |\mathbf{x}_t\rangle \, \big]$.

To use the tICA method, we must construct estimators for μ, Σ, 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 τ, $\left\{\left(|X_t\rangle, |Y_t\rangle\right)\right\}_{t=1}^{M}$. 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:

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

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:

$$\log L = \sum_{t=1}^{M} \left[-\frac{d}{2}\log 2\pi - \frac{1}{2}\log|S| - \frac{1}{2}\Big(\langle Z_t| - \langle m|\Big) S^{-1} \Big(|Z_t\rangle - |m\rangle\Big)\right] \tag{3}$$

Using the properties of the matrix differential, it can be shown that the total differential of the log-likelihood is:

$$
\begin{aligned}
d\log L = &-\frac{M}{2}\,\mathrm{tr}(S^{-1}dS) \\
&-\frac{1}{2}\sum_{t=1}^{M}\mathrm{tr}\left[2S^{-1}\Big(\left|Z_t\right\rangle - \left|m\right\rangle\Big)d\left\langle m\right|\right] \\
&-\frac{1}{2}\sum_{t=1}^{M}\mathrm{tr}\left[S^{-1}\Big(\left|Z_t\right\rangle - \left|m\right\rangle\Big)\Big(\left\langle Z_t\right| - \left\langle m\right|\Big)S^{-1}dS\right]
\end{aligned} \tag{4}
$$

We could rewrite this total differential in terms of $|\mu\rangle$, $C^{(\tau)}$, and Σ, 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:

$$
RSR = S \qquad \text{and} \qquad R\left|m\right\rangle = \left|m\right\rangle.
$$

We can then construct the Lagrange function:

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

The total differential of Λ can then be written:

$$
\begin{aligned}
d\Lambda = \operatorname{tr}\Bigg[\Bigg(-\sum_{t=1}^{M} S^{-1}\Big(\left|Z_t\right\rangle - \left|m\right\rangle\Big) + (R-I)\left|\lambda\right\rangle\Bigg) d\left\langle m\right|\Bigg] \\
+ \operatorname{tr}\left[\left(-\frac{M}{2}S^{-1} - \frac{1}{2}\sum_{t=1}^{M} S^{-1}\Big(\left|Z_t\right\rangle - \left|m\right\rangle\Big)\Big(\left\langle Z_t\right| - \left\langle m\right|\Big)S^{-1} + (R\Phi R - \Phi)\right) dS\right] \\
+ \operatorname{tr}\Big[\Big(R\left|m\right\rangle - \left|m\right\rangle\Big) d\left\langle \lambda\right|\Big] + \operatorname{tr}\Big[\Big(RSR - S\Big)d\Phi\Big]
\end{aligned} \tag{5}
$$

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

$$
\begin{aligned}
-\sum_{t=1}^{M} S^{-1}\Big(\left|Z_t\right\rangle - \left|m\right\rangle\Big) + (R-I)\left|\lambda\right\rangle = 0 \\
\left(\sum_{t=1}^{M}\left|Z_t\right\rangle\right) - M\left|m\right\rangle = S(R-I)\left|\lambda\right\rangle
\end{aligned}
$$

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

$$
\begin{aligned}
\left[\left(\sum_{t=1}^{M}\left|Z_t\right\rangle\right) - M\left|m\right\rangle\right] + R\left[\left(\sum_{t=1}^{M}\left|Z_t\right\rangle\right) - M\left|m\right\rangle\right] &= \Big[S(R-I)\left|\lambda\right\rangle\Big] + R\Big[S(R-I)\left|\lambda\right\rangle\Big] \\
\left(\sum_{t=1}^{M}\left|Z_t\right\rangle + R\left|Z_t\right\rangle\right) - 2M\left|m\right\rangle &= (SR - S + RSR - RS)\left|\lambda\right\rangle \\
\left(\sum_{t=1}^{M}\left|Z_t\right\rangle + R\left|Z_t\right\rangle\right) - 2M\left|m\right\rangle &= \Big[(SR - RS) + (RSR - S)\Big]\left|\lambda\right\rangle \\
\left(\sum_{t=1}^{M}\left|Z_t\right\rangle + R\left|Z_t\right\rangle\right) - 2M\left|m\right\rangle &= 0 \\
\left|m\right\rangle = \frac{1}{2M}\sum_{t=1}^{M}\left|Z_t\right\rangle + R\left|Z_t\right\rangle
\end{aligned}
$$

We can then solve for S.

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

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

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

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

$$= S(R\Phi R - \Phi)S + R\left(S(R\Phi R - \Phi)S\right)R$$

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

$$= SR\Phi RS - S\Phi S + RSR\Phi RSR - RS\Phi SR$$

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

$$= \left(SR\Phi RS - RS\Phi SR\right) + \left(RSR\Phi RSR - S\Phi S\right)$$

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

$$= 0$$

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

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

$$\left|\mu_{\text{mle}}\right\rangle = \frac{1}{2N}\sum_{t=1}^{N}\left(\left|\delta X_t\right\rangle + \left|\delta Y_t\right\rangle\right) \tag{6}$$

$$\Sigma_{\text{mle}} = \frac{1}{2N}\sum_{t=1}^{N}\left(\left|\delta X_t\right\rangle\left\langle\delta X_t\right| + \left|\delta Y_t\right\rangle\left\langle\delta Y_t\right|\right) \tag{7}$$

$$C^{(\tau)}_{\text{mle}} = \frac{1}{2N} \sum_{t=1}^{N} \Big(|\delta X_t\rangle \langle \delta Y_t| + |\delta Y_t\rangle \langle \delta X_t| \Big) \tag{8}$$

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:

$$\frac{\langle v| \, C^{(\tau)}_{\text{mle}} \, |v\rangle}{\langle v| \, \Sigma_{\text{mle}} \, |v\rangle}$$

is always in $[-1, 1]$ (as long as Σ_{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 Σ is positive definite, it is also nonsingular and so we can write:

$$
\begin{aligned}
|v\rangle &= \frac{1}{\lambda}\Sigma^{-1}C^{(\tau)}|v\rangle \\
&= \frac{1}{2M\lambda}\Sigma^{-1}\sum_{t=1}^{M}\Big(\langle\Phi(X_t)|\,v\rangle\Big)|\Phi(Y_t)\rangle + \Big(\langle\Phi(Y_t)|\,v\rangle\Big)|\Phi(X_t)\rangle \\
&:= \frac{1}{2M\lambda}\Sigma^{-1}|x\rangle
\end{aligned}
$$

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:

$$
\Sigma = P\Lambda P^T = \Big[\,|p_1\rangle\,|p_2\rangle\ldots|p_d\rangle\,\Big]\begin{bmatrix}\lambda_1 & 0 & \ldots & 0\\ 0 & \lambda_2 & 0 & \ldots\\ & & \vdots & \\ 0 & 0 & \ldots & \lambda_d\end{bmatrix}\begin{bmatrix}\langle p_1|\\ \langle p_2|\\ \vdots\\ \langle p_d|\end{bmatrix}
$$

where $|p_i\rangle$ is an eigenvector of Σ and λ_i is its eigenvalue. It's easy to show that the eigenvectors of Σ 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 Σ:

$$
\begin{aligned}
|v\rangle &= \frac{1}{2M\lambda}P\Lambda^{-1}P^T|x\rangle \\
&= \frac{1}{2M\lambda}\Big[\,|p_1\rangle\,|p_2\rangle\ldots|p_d\rangle\,\Big]\Lambda^{-1}\begin{bmatrix}\langle p_1|\,x\rangle\\ \langle p_2|\,x\rangle\\ \vdots\\ \langle p_d|\,x\rangle\end{bmatrix} \\
&= \frac{1}{2M\lambda}\sum_{i=1}^{d}\Big(\lambda_i^{-1}\langle p_i|\,x\rangle\Big)|p_i\rangle
\end{aligned}
$$

which means that $|v\rangle$ is in the span of the eigenvectors of Σ. Since the eigenvectors of Σ 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).

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

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

$$|v\rangle = \sum_{t=1}^{M} \beta_i\, |\Phi(X_t)\rangle + \beta_{i+M}\, |\Phi(Y_t)\rangle \tag{10}$$

Now, we need to simply expand the numerate and denominator of Eq. (9) in terms of the elements of β.

$$
\begin{aligned}
\langle v| C^{(\tau)} |v\rangle &= \frac{1}{2M}\sum_{t=1}^{M} \langle v|\, \Phi(X_t)\rangle \langle \Phi(Y_t)|\, v\rangle \\
&\qquad + \langle v|\, \Phi(Y_t)\rangle \langle \Phi(X_t)|\, v\rangle \\
&= \frac{1}{2M}\sum_{t=1}^{M}\Bigg(\sum_{i=1}^{M}\left[\beta_i K_{it}^{XX} + \beta_{i+M}K_{it}^{YX}\right] \sum_{j=1}^{M}\left[\beta_j K_{tj}^{YX} + \beta_{j+M}K_{tj}^{YY}\right] \\
&\qquad + \sum_{i=1}^{M}\left[\beta_i K_{it}^{XY} + \beta_{i+M}K_{it}^{YY}\right] \sum_{j=1}^{M}\left[\beta_j K_{tj}^{XX} + \beta_{j+M}K_{tj}^{XY}\right]\Bigg) \\
&= \frac{1}{2M}\sum_{t=1}^{M}\left[\beta^T \begin{pmatrix} K^{XX} \\ K^{YX} \end{pmatrix}\right]_t \left[\begin{pmatrix} K^{YX} & K^{YY} \end{pmatrix}\beta\right]_t \\
&\qquad + \left[\beta^T \begin{pmatrix} K^{XY} \\ K^{YY} \end{pmatrix}\right]_t \left[\begin{pmatrix} K^{XX} & K^{XY} \end{pmatrix}\beta\right]_t \\
&= \frac{1}{2M}\beta^T \left[\begin{pmatrix} K^{XX} \\ K^{YX} \end{pmatrix}\begin{pmatrix} K^{YX} & K^{YY} \end{pmatrix} + \begin{pmatrix} K^{XY} \\ K^{YY} \end{pmatrix}\begin{pmatrix} K^{XX} & K^{XY} \end{pmatrix}\right]\beta \\
&= \frac{1}{2M}\beta^T \begin{bmatrix} K^{XX}K^{YX} + K^{XY}K^{XX} & K^{XX}K^{YY} + K^{XY}K^{XY} \\ K^{YX}K^{YX} + K^{YY}K^{XX} & K^{YX}K^{YY} + K^{YY}K^{XY} \end{bmatrix}\beta \\
&= \frac{1}{2M}\beta^T \begin{bmatrix} K^{XX} & K^{XY} \\ K^{YX} & K^{YY} \end{bmatrix}\begin{bmatrix} K^{YX} & K^{YY} \\ K^{XX} & K^{XY} \end{bmatrix}\beta \\
&= \frac{1}{2M}\beta^T K \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix} K\beta
\end{aligned}
$$

Let $R = \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix}$, then the numerator becomes:

$$\langle v| \, C^{(\tau)} \, |v\rangle = \frac{\beta^T K R K \beta}{2M} \tag{11}$$

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

$$\langle v| \, \Sigma \, |v\rangle = \frac{\beta^T K K \beta}{2M} \tag{12}$$

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. $\mathbb{E}\big[\, |\Phi(X_t)\rangle \,\big] = 0$). 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:

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

where $\mathbf{1}$ 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.